

A Grammar of Kharia

Brill's Studies in South and Southwest Asian Languages

Series Editors

John Peterson, University of Leipzig
Anju Saxena, Uppsala University (editor of this volume)

Editorial Board

Anvita Abbi, Jawaharlal Nehru University
Balthasar Bickel, University of Leipzig
George Cardona, University of Pennsylvania
Carol Genetti, University of California, Santa Barbara
Geoffrey Haig, University of Bamberg
Gilbert Lazard, CNRS & École pratique des Hautes Études
Harold Schiffman, University of Pennsylvania
Udaya Narayana Singh, Visva-Bharati, Shantiniketan, India

VOLUME 1

A Grammar of Kharia

A South Munda Language

By

John Peterson



BRILL

LEIDEN • BOSTON
2011

On the cover

Front: This photograph shows the hill (Kharia *biru*) which gave the village of Biru located at the foot of this hill in southwestern Jharkhand, near the city of Simdega, its name. This village is said to have been the traditional home of the Kharia “kings” (*kolen*). The present-day “king’s” family hails from Orissa and is not Kharia.

Back: This photograph shows the stone formation referred to locally as the “throne”, located on this same hill (not visible on the front cover) and is undoubtedly the reason that this small village was chosen by earlier rulers as their capitol.

This book is printed on acid-free paper.

Library of Congress Cataloging-in-Publication Data

Peterson, John M. (John Michael)

A grammar of Kharia a South Munda language / by John Peterson.

p. cm. — (Brill’s studies in South and Southwest Asian languages)

Text in English and Kharia.

Includes bibliographical references and index.

ISBN 978-90-04-18720-7 (acid-free paper) 1. Kharia language—Grammar.

I. Title. II. Series.

PL4579.P48 2011

495.9’5—dc22

2010043092

ISSN 1877-4083

ISBN 978 90 04 18720 7

Copyright 2011 by Koninklijke Brill NV, Leiden, The Netherlands.

Koninklijke Brill NV incorporates the imprints Brill, Hotei Publishing,

IDC Publishers, Martinus Nijhoff Publishers and VSP.

All rights reserved. No part of this publication may be reproduced, translated, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without prior written permission from the publisher.

Authorization to photocopy items for internal or personal use is granted by Koninklijke Brill NV provided that the appropriate fees are paid directly to The Copyright Clearance Center, 222 Rosewood Drive, Suite 910, Danvers, MA 01923, USA. Fees are subject to change.

In memory of

Mr. Stanislas Kullu,

previous head of the All-India Kharia Society, an active supporter of Kharia, a highly respected and active member of his community, and a very dear friend.

for Marco

CONTENTS

List of Maps	xv
Acknowledgements	xvii
Abbreviations and Symbols	xxi
Abbreviations of Sources	xxiii
Chapter One Introduction: The Kharia Language	1
1.1 General Introduction	1
1.2 Genetic Affiliations	7
1.3 Previous Studies on the Kharia and What this Study Adds	8
1.3.1 Previous works	8
1.3.2 The present study	11
1.3.3 Methodology	13
1.4 Some Terminological Issues	16
1.4.1 “Contentive morphemes and semantic bases” vs. “roots and stems”	17
1.4.2 “Case- and TAM/PERSON-syntagmas” vs. “nouns, verbs and adjectives”	18
1.4.3 “Marker of qualitative predication” vs. “copula”	19
1.5 Kharia Orthography and the System of Transliteration Used here	19
1.5.1 The script	20
1.5.2 The written word in Kharia	24
1.6 The Structure of this Study	24
Chapter Two Phonology	27
2.1 Vowel Inventory	27
2.1.1 Monophthongs	27
2.1.2 Diphthongs	28
2.2 Consonants	29
2.2.1 Gemination	31
2.2.2 Pre-glottalized stops	31
2.3 Syllable Structure	32
2.4 Morphophonology	33
2.5 The Phonological Word	35
2.6 Sentence Prosody	46

Chapter Three	Syntactic Atoms	51
3.1	Clitics as “Phrasal Affixes”	51
3.1.1	The pragmatic markers	53
3.1.2	Case and number marking	55
3.1.3	Markers for inalienable possession	57
3.1.4	Markers for TAM and person / number / honorific status	58
3.1.5	The perfect marker = <i>si?</i> (<i>q</i>) and the infinitive marker = <i>na</i>	60
3.2	Clitics in Kharia—An Overview of their Characteristics	61
3.2.1	Position with respect to host	61
3.2.2	Segmental properties of clitics and their respective degree of cohesion	62
3.2.2.1	Segmental properties	62
3.2.2.2	Degree of cohesion with host	63
3.2.3	Clitics as “reduced” phonological words	65
3.2.4	The relative ordering of clitics	65
3.2.5	The varying mobility of clitics	67
3.2.6	Lexicalization	68
3.3	Affixes	68
Chapter Four	Parts of Speech and the Lexicon	71
4.1	Introduction and Methodology	71
4.2	The “Precategoriality” of Simple Contentive Morphemes	75
4.3	Derived Forms	92
4.3.1	Bisyllabicity in Kharia	93
4.3.2	Reduplication—The masdar	94
4.3.3	The <i>-NV-</i> infix	101
4.3.4	<i>-NV-</i> + Reduplication	104
4.3.5	Derivation and parts of speech—A summary	106
4.4	“NPs” as Verbs? Syntactic Categories, not Lexical Categories	106
4.4.1	The question of acceptability	112
4.5	Parts of Speech in Kharia: A Summary	116
4.6	“Compounds” and “Noun Incorporation”	118
4.6.1	Lexical juxtapositions as “compounds”	118
4.6.2	Genuine compounds	121
4.6.2.1	(Semi-)Productive compounding	123
4.6.3	Pseudo-compounds and “noun incorporation”	124
4.6.4	Semi-productive incorporation	128
4.7	“Echo-words”	129

Chapter Five	The Case-syntagma (“NPs”)	135
5.1	General Introduction	135
5.1.1	Gender / size and animacy as grammatical categories	139
5.1.2	Definiteness and indefiniteness as grammatical categories	142
5.2	Case	143
5.3	Adpositions	149
5.4	Number	158
5.5	Inalienable Possession	162
5.6	Proforms	168
5.7	Interrogatives, Indefinites and Negative Indefinites	178
5.8	Demonstratives	182
5.9	Quantifiers and Classifiers	187
5.9.1	Quantifiers	187
5.9.2	Classifiers	194
5.10	Modification	200
5.10.1	Morphologically simple modification (“adjectives”)	200
5.10.1.1	Comparatives, superlatives and equatives	202
5.10.2	Clausal adjuncts (“adverbials”)	205
Chapter Six	The TAM/PERSON-syntagma (“Verbs”)	209
6.1	General Introduction	209
6.2	Person / Number / Honorific Marking	212
6.3	Types of semantic bases in the TAM/PERSON-syntagma	214
6.3.1	Single contentive morpheme as the semantic base	214
6.3.2	TAM/PERSON-syntagmas with multiple semantic sub-bases	215
6.3.3	Other types of semantic bases	219
6.3.4	Experiential predicates / “psyche verbs”	226
6.3.4.1	Experiencer as subject	226
6.3.4.2	Oblique-case experiencers	227
6.3.4.3	Overview of experiential predicates	228
6.3.5	Causative and reciprocal marking	230
6.4	TAM / Basic Voice Marking	239
6.4.1	The basic TAM categories	239
6.4.1.1	Past and “Past II”	240
6.4.1.2	Present general imperfective	243

6.4.1.3	Present progressive	245
6.4.1.4	Irrealis	247
6.4.1.5	Perfect	249
6.4.1.6	Optative	257
6.4.2	Active and Middle: Their Functions and Distribution	258
6.4.2.1	General remarks	258
6.4.2.2	Differential functions	260
6.4.2.3	Inherent functions	276
6.4.2.4	A semantic map of the middle voice in Kharia	288
6.5	“V2s”—Markers of the Passive / Reflexive and <i>Aktionsart</i>	293
6.5.1	Passive / Reflexive <i>qom</i>	293
6.5.2	“Durativity”—continuative <i>kan</i> , semel-iterative <i>lo?</i> and iterative <i>khor</i>	298
6.5.3	Anticipatory telic <i>qoʔq</i> and culminatory telic <i>goʔq</i>	302
6.5.4	Totality <i>may</i>	304
6.5.5	Ambulative <i>san</i> ‘while going’	306
6.5.6	Suddenness <i>bha?</i> , <i>hamba?</i> and <i>dhaʔb</i>	306
6.5.7	Departive <i>tu</i>	307
6.5.8	Conative <i>dakha</i> / <i>lakha</i>	308
6.5.9	Excessive <i>bay</i>	308
6.5.10	Autopoesis <i>jom</i>	309
6.5.11	Benefactive <i>kay</i>	309
6.5.12	Seldomly encountered V2s	310
6.5.12.1	<i>col</i> ‘go’	310
6.5.12.2	<i>oʔj</i> ‘take out’	310
6.5.12.3	<i>ol</i> ‘take’	311
6.5.12.4	<i>pal</i> ‘finish’	311
6.5.12.5	<i>saphay</i> ‘clean’	312
6.6	Partially Finite and Non-finite Forms	313
6.6.1	Partially finite forms	313
6.6.2	Non-finite forms	317
6.6.2.1	The masdar	317
6.6.2.2	Infinitives	322
6.6.2.3	Sequential converbs (“conjunctive participles”)	324
6.6.2.4	Imperfective converbs (“present participles”)	330
6.6.2.5	Participles	333

6.7	Negation	334
6.7.1	The general negator <i>um</i>	335
6.7.2	Modal negation with <i>abu</i>	338
6.7.3	Emphatic negation with <i>umbo?</i>	339
6.7.4	Negation with <i>na</i>	341
6.8	Periphrastic TAM/PERSON-syntagmas (“Verbs”)	341
6.8.1	“Aspectual” auxiliaries	342
6.8.1.1	<i>la?</i> ‘imperfectivity’	342
6.8.1.2	<i>māre/mare, suru</i> ‘begin’	345
6.8.1.3	<i>melay</i> ‘stop’	346
6.8.1.4	<i>col</i> ‘go’	346
6.8.2	Modality	346
6.8.2.1	<i>lam</i> ‘look for; want’	346
6.8.2.2	<i>ter</i> ‘give; allow’	347
6.8.2.3	<i>pal</i> ‘can, be able’	348
6.8.2.4	The qualitative predicative marker: Obligation and necessity	348
6.9	Inference	349
Chapter Seven	Syntax	351
7.1	Grammatical Relations	351
7.1.1	Case-syntagmas and their status within the clause	352
7.1.2	Subject	359
7.1.3	Object types	367
7.2	Interrogatives	368
7.2.1	Polar questions	369
7.2.2	Constituent (“wh-”) questions	370
7.2.2.1	Multiple constituent questions	371
7.2.3	Alternative questions	373
7.3	Qualitative Predication—“Nominal” and “Copular” Sentences	374
7.3.1	General introduction	374
7.3.2	Does Kharia have a copula?	375
7.3.3	Stative qualitative predication—the forms	377
7.3.4	Dynamic qualitative predication	381
7.3.5	“Zero copulas”?	382
7.4	Clause Coordination	388
7.5	Subordination	389
7.5.1	“Adverbial” clauses	389
7.5.1.1	Purpose clauses	389
7.5.1.2	Causal / instrumental clauses	390

7.5.1.3	Temporal clauses	390
7.5.1.4	Conditionals and counterfactuals	393
7.5.1.5	Concessives	395
7.5.1.6	Depictives	396
7.5.2	Complement clauses	403
7.5.2.1	In subject function	403
7.5.2.2	In object function	404
7.6	“Relative Clauses”—Propositional Attribution	406
7.6.1	Correlatives	408
7.6.2	Prenominal constructions	410
7.6.3	Circumnominal clauses	419
7.6.4	Postnominals	421
7.6.5	The scale of relativization	422
7.7	Information Structure—Topic, Focus and Constituent Order	426
7.7.1	Predicate focus or Topic-Comment	427
7.7.2	Argument focus	432
7.7.3	Sentence focus	432
7.7.4	The pragmatic markers = <i>ga</i> , = <i>ko</i> and = <i>jo</i>	433
Appendix	Text: The Nine Totems	439
	Bibliography	453
	General Index	461
	Index of Authors	473

LIST OF MAPS

Map 1: The States of Jharkhand, Chattisgarh and Orissa in eastern central India	2
Map 2: The districts where Kharia is spoken	3

ACKNOWLEDGEMENTS

The present study is a thoroughly revised version of the first volume of my *Habilitationsschrift* or “professorial dissertation” Volume I of that study was primarily a reference grammar of Kharia and was itself a greatly expanded version of a shorter grammar which appeared in a multi-contributor work, *The Munda Languages*, edited by Gregory D.S. Anderson (Routledge, 2008, ISBN 978-0415-32890-6). I gratefully acknowledge here my gratitude to Routledge for granting permission to re-publish this earlier, shorter grammar in this revised and greatly expanded form.

The remaining two volumes of that study included a collection of Kharia texts (Volume II), segmented, glossed, annotated and translated into English, which is now freely accessible.¹ Volume III of that study was a Kharia-English lexicon which has since been published online as Peterson (2009).²

As with any work of this type, many people have greatly aided me in my work, and without their help this volume would never have been possible. I would like to take this opportunity to thank these individuals, and I apologize in advance if I have inadvertently overlooked anyone here.

First of all, my work on Kharia was greatly assisted by two generous grants from the German Research Foundation (*Deutsche Forschungsgemeinschaft*, DFG) (PE 872/1-1, 2), which I gratefully acknowledge here.

I would also like to thank the many individuals in Jharkhand who made my research trips there both pleasant and productive. First of all, thanks are due to those who taught me Kharia, who bore with me while I was still stuttering and searching for words, and who spent hours upon hours discussing the intricacies of their native language with me: Mr. Basil Baa, Ms. Rayem Olem Baa (formerly Dungdung), Ms. Rose Dungdung, Ms. Saroj Kerketta, Mr. Silvester Kerketta, Mr. Anugrah Kullu, Ms. Tarke-leng Kullu, Ms. Anita Soreng, Mr. Dilip Soreng, Mr. Marcus Soreng, Mr. Pratap Soreng, and Mr. Marianus Tete. It has been a great pleasure for me to get to know you all and to be able to work with you!

¹ http://www.southasiabibliography.de/Bibliography/Austroasiatic/Munda/Kharia/Kharia_A_South_Munda_Language/kharia_a_south_munda_language.html

² <http://www.linguistics.ucsb.edu/HimalayanLinguistics/grammars/HLA05.html>

Thanks also to the “Kharia Boys” of St. Alois Minor Seminary in Ranchi: Sanjay Dung Dung, Edwin R. Dung Dung, Ajit Kerketta, Alexzendar Kerketta, Pavan Prakash Kerketta, Ravi Kujur, Ranjit Kullu, Samir Kullu, Vinay Kullu, Alok Soreng, Anil Soreng, Shilanan Tete, and of course our “Kurukh supporter” Pascal Toppo, for taking the time in the evening after a long and busy day to answer my questions and for seeing to it that the Seminary’s watchdog Tommy didn’t bite me again!

My sincerest thanks also go to Mr. Anil Birendra Kullu, the first Kharia speaker I met and worked with, and to his father, the late Mr. Stanislas Kullu, who welcomed me into his home and made me feel a part of his family, and who accompanied me to Birmitrapur and introduced me there to Mr. Marcus Soreng. Thanks also to the wonderful people of Saldega (*salḍa?*), who made me feel welcome in their community and who greatly assisted me with my work. I am also grateful to Father Soreng of the Simdega Development Centre (*simḍegā vikās kendra*) for all the pleasant conversations, the hospitality he and his fellow brethren showed me, despite their busy schedules, and for providing me with a place to stay in Simdega on one occasion when I had nowhere else to turn.

Thanks also go to the wonderful students and faculty of the Department of Tribal and Regional Languages at Ranchi University for all their hospitality, especially Ms. Mary Soreng for her kindness, her insights into Kharia grammar, and for introducing me to the many Kharia scholars of Ranchi; to Ms. Rose Kerketta, who has probably published more in (and on) Kharia than anyone else, for her hospitality and for sharing her vast knowledge of Kharia with me; and finally to Louni Kandulna, for all her kindness and help.

I am also grateful to Fathers Joachim Dungdung and Nicholas Tete, two renowned scholars of the Kharia and their language, both of whom took an immediate interest in my work, for discussing my work with me, for showing me their own work on Kharia, for their many valuable suggestions, and of course for all the wonderful conversations.

Thanks also to Dr. Ram Dayal Munda, previous head of the Department of Tribal and Regional Languages and now an important political figure in Jharkhand, for supporting me wherever he could with my work, despite his extremely tight schedule, often taking time between his interviews with reporters to talk with me.

And finally, very special thanks go to Dr. Ganesh Murmu. Although he himself is not a Kharia speaker, he has aided me in my work perhaps more than anyone else. It was he who first welcomed me to Ranchi, although we had never even met, who found me a place to stay, showed me around

the Department, introduced me to speakers of Kharia, and helped pave the way for my work. There are few people who know Jharkhand as well as he does, and I cannot imagine that my work in Jharkhand would ever have been successful were it not for him.

Outside of Jharkhand, there are also many people who have contributed to this work in some way, and I apologize in advance if I have forgotten anyone. First of all, the present study has benefitted greatly from my discussions with participants at a number of workshops, conferences and talks at the following universities, and I am grateful for all the constructive feedback and suggestions that I have received from everyone there: Leipzig (Max Planck Institute, University of Leipzig), the universities of Bochum, Osnabrück, Lund, Uppsala, Konstanz, Göttingen and Hamburg, Delhi University and Jawaharlal Nehru University in New Delhi. Further thanks go to Anvita Abbi, K.V Subbarao, Tatiana Oranskaia, Dörte Borchers, Olaf Müller-Reichau and Anju Saxena for inviting me to speak at their departments and for all the stimulating discussions which followed. Anju Saxena was also solely responsible for processing the present volume, making sure that it was subjected to the same stringent scrutiny as any other manuscript in this series (including getting the manuscript double-blind reviewed).

Thanks also go to Gregory D.S. Anderson, Walter Bisang, Anne Daladier, Matthew Dryer, Christoph Gabriel, Arlo Giffiths, Graham Katz, Ulrich Mehlem, Toshiki Osada, Christoph Schroeder, David Stampe, and two anonymous reviewers for comments and suggestions on the many, many, many previous versions of this work and other related articles which helped pave the way for the present study.

And finally, special thanks go to Utz Maas for holding down the fort at the department in Osnabrück while I spent several months in India gathering data and for his constant suggestions and constructive criticism of my work, forcing me at every step to re-consider my argumentation.

Needless to say, I alone am of course responsible for any and all oversights and errors which the present study may contain.

ampe sou²bte baruḍa? gamtīn! iku²d.jughay dheinbad!

ABBREVIATIONS AND SYMBOLS

SYMBOLS

- < > infix marker; also used to mark a written (as opposed to spoken) word or symbol
- denotes a derivational affix or compounded element
- = denotes an enclitic

ABBREVIATIONS

Numbers in parentheses refer to the respective section(s) where these topics are discussed in more detail.

- | | |
|--------------------------------------|--|
| ABL—ablative (5.2) | ECHO—echo-word (4.7) |
| ACT—active (6.4.2) | EMOT—emotive predicate marker (6.3.4) |
| ADD—additive focus (7.7.4) | EXCL—exclusive (5.4; 6.2) |
| AMB—ambulative (6.5.5) | EXCES—excessive (6.5.9) |
| ANAPH—anaphoric proform (5.6) | FOC—restrictive focus; also used for various other focus markers whose exact function awaits further study (<i>ro?</i> , <i>ga?</i> , etc.) (7.7.4) |
| APPROX—approximative (3.3) | GEN—genitive (5.2) |
| A:TEL—anticipatory telic (6.5.3) | HON—honorific (5.4; 6.2) |
| AUTOPOES—autopoesis (6.5.10) | HUM—human (5.6) |
| BEN—benefactive (6.5.11) | IA—Indo-Aryan |
| CAUS—causative (6.3.5) | INCL—inclusive (5.4; 6.2) |
| CLASS—numeral classifier (5.9.2) | INDEF—indefinite |
| CMPL—complementizer | INF—infinite (6.6.2.2) |
| CNTR—contrastive focus (7.7.4) | INFER—inferential (6.9) |
| CONAT—conative (6.5.8) | INST—instrumental (5.3) |
| CONT—continuative (6.5.2) | INTENS—intensifier (5.10.2) |
| CREL—correlative (7.6.1) | IPFV—imperfective (6.8.1.1) |
| C:TEL—culminatory telic (6.5.3) | IRR—irrealis (6.4.1.4) |
| CVB—(imperfective) converb (6.6.2.4) | ITER—iterative (6.5.2) |
| DU—dual (5.4; 6.2) | |
| DEM—demonstrative (5.8) | |
| DIST—distal demonstrative (5.8) | |
| DPT—departive (6.5.7) | |

- ITR—intransitive
 Kh.—Kharia
 LOC—locative *hinte* (5.3)
 MID—middle (6.4.2)
 MOD—modal (negation) (6.7.2)
 -*m*- hiatus-breaking glide
 NHUM—non-human (5.6)
 OBL—oblique (5.2)
 OPT—optative (6.4.1.6)
 PASS—passive (6.5.1)
 PERF—perfect (6.4.1.5)
 PL—plural (5.4; 6.2)
 POSS—inalienable possession
 (5.5)
 PRS—present tense (6.4.1.2)
 PROG—progressive (6.4.1.3)
 PST—past tense (6.4.1.1)
 PST.II—“Past II” (6.4.1.1)
 PTCP—participle (6.6.2.5)
 PURP—purposive (7.5.1.1)
 Q—interrogative (7.2.1)
 QUAL—qualitative predication
 (7.3)
- RDP—reduplicated form
 (= *masdar*, 4.3.2; 6.6.2.1). Only
 the stem is reduplicated.
 REC—reciprocal (6.3.5)
 REFL—reflexive (6.5.1)
 REP—repetition (with converbs,
 intensification, “echo-words”,
 plurality). Repetition of an
 entire phonological word.
 SEQ—sequential converb (6.6.2.3)
 SG—singular (5.4; 6.2)
 S:ITER—semel-iterative (6.5.2)
 SUD—suddenness (6.5.6)
 TEMP—temporal (“relative
 clause”) (7.6)
 TOTAL—totality (6.5.4)
 TR—transitive
 V2—*Aktionsart* marker or “v2”
 following semantic base of
 predicate and preceding TAM
 markers (6.5)
 VOC—vocative
 -*w*-, -*y*- hiatus-breaking glides

ABBREVIATIONS OF SOURCES

- BG Biligiri, H.S. 1965. *Kharia: Phonology, grammar and vocabulary*. Poona: Deccan College, Postgraduate and Research Institute. (Building Centenary and Silver Jubilee Series, 3).
- HJPa Pinnow, Heinz-Jürgen. 1965a. *Kharia-Texte (Prosa und Poesie)*. Wiesbaden: Harrassowitz.
- HJPb Pinnow, Heinz-Jürgen. 1965b. “Der Wertiger” und andere Geschichten in Kharia, aufgezeichnet von Junas Kullu, übersetzt und mitgeteilt unter Mitwirkung von Paulus Kerketta.” *Indo-Iranian Journal* 9/1: 32–68.
- RR Roy, Sarat Chandra & Ramesh Chandra Roy. 1937. *The Khārīās*. Ranci: Man in India.

Sources of the type “[AK, 1:2]” refer to examples from my own corpus. These examples are abbreviated as follows: E.g., in [AK, 1:2], “AK” refers to the speaker and “1:2” to the text and text-line in my corpus, respectively. Thus, [AK, 1:2] refers to the second line of the first text by the speaker AK.

Full-length examples for which no source is given are from my own interviews and conversations. All other examples are cited in the usual format.

INTRODUCTION: THE KHARIA LANGUAGE

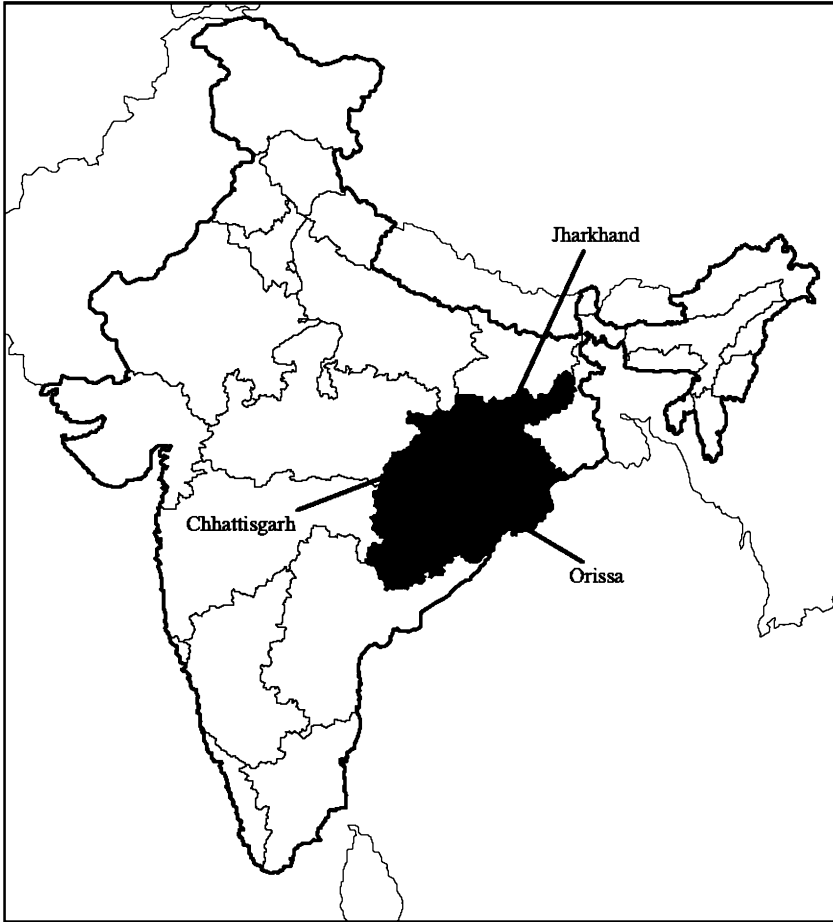
1.1 GENERAL INTRODUCTION

Kharia is a South Munda language spoken primarily in eastern central India in the southwestern areas of the state of Jharkhand (mainly in the districts Simdega and Gumla) and adjoining areas of the states of Chattisgarh (primarily in the districts Surguja, Jashpur and Raigarh) and Orissa (especially in Sundargarh District). It is also spoken by large numbers of speakers in regions further to the east and south, as well as by smaller numbers further afield (cf. the discussions and map in Ishtiaq, 1999: 41ff.). The largest number of speakers probably reside in the Simdega and Gumla districts of Jharkhand. Map 1 shows the location of the states of Jharkhand, Chattisgarh and Orissa (shaded) within India while Map 2 shows the districts where Kharia is predominantly spoken in these states. According to Lewis (2009),¹ there is also a considerable number of speakers in Assam and smaller communities elsewhere, including West Bengal and the Andaman and Nicobar Islands. The language described in the present study is that spoken in Simdega and Gumla Districts, Jharkhand, as well as in Birmitrapur, on the Orissa side of the Jharkhand-Orissa border.

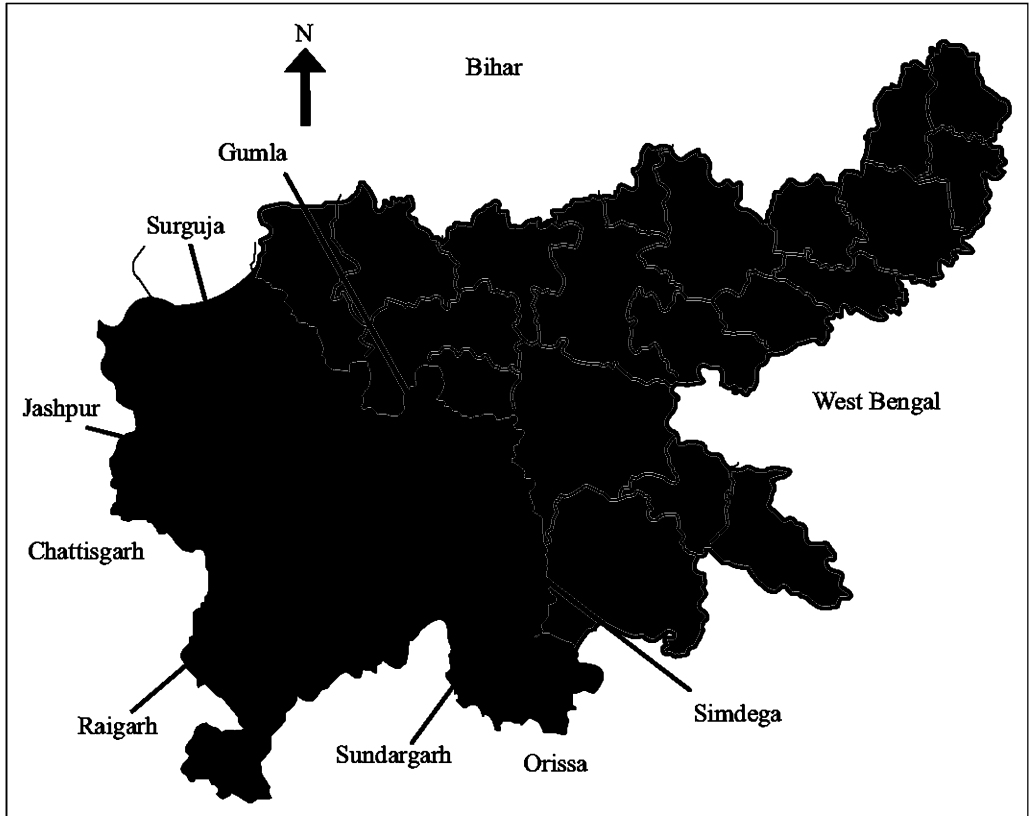
Kharia is a predominantly agglutinating language, and virtually all grammatical marking is expressed through enclitics, although there is also a limited number of derivational suffixes, one prefix and two infixes, one of which appears either as a prefix or infix. The language is generally predicate-final, although not rigidly so, and the order of the Case-syntagmas (“NPs”) is determined primarily through their pragmatic status, with topical elements appearing first, followed by focal elements. There are only two overtly expressed cases, the genitive and the oblique, the latter being used to mark a variety of argument-types, including (specific and definite) direct objects, “indirect objects”, as well as adjuncts of various types.

Perhaps the most striking characteristic of the language is the virtually total lack of evidence for lexical categories such as noun, verb and adjective,

¹ http://www.ethnologue.com/show_language.asp?code=KHR



Map 1: The States of Jharkhand, Chattisgarh and Orissa in eastern central India.



Map 2: The districts where Kharia is predominantly spoken.

at least in the Simdega dialect, and predicates (\approx Vs) and their complements (\approx NPs) are in fact based on syntactic structures, not lexical units.

Typologically speaking, Kharia resembles a large number of other South Asian languages in many respects and can probably be considered as belonging to what is often termed the South Asian linguistic area or *Sprachbund*, although a precise characterization of this area has so far eluded researchers. For example, Kharia possesses all five of the criteria discussed at length in Masica (1976) as typical of the South Asian linguistic area, namely head-final order, morphologically marked causatives (and double causatives), the presence of sequential converbs (or “conjunctive participles”), explicator verbal auxiliaries, referred to here as “v2s”, i.e., markers of *Aktionsart* which derive from earlier contentive morphemes (usually verbs in languages for which this class may be assumed to exist), and “dative subjects”, i.e., constructions in which the experiencer appears in a case other than the nominative (or the direct case, as it is termed here) and in which the stimulus is the grammatical subject of the construction.

However, as Masica himself notes, these five criteria are to a certain extent rather arbitrary, and it could well be argued that at least most of them are nothing out of the ordinary for a predicate-final language. In fact, a number of other researchers have proposed numerous possible criteria over the past several decades, many of which are summarized in Masica (1976, Appendix A), going back at least to Emeneau’s (1956) seminal study.

More recent studies are increasingly taking data on languages into account which have only become available in the past few decades, especially on languages of the east and northeast. Some of the more recent studies in this area include Ebert, 1993; 1998; 1999; Ebert & Neukom, 2000; Moral, 1997; Neukom, 1999. Another innovative study in the field of contact linguistics is Donegan & Stampe (2004), in which the authors argue that much of the common structure of the Munda and Indo-Aryan languages can be traced back to prosodic features. It is hoped that the present study will, among other things, prove useful in future studies in this field, especially since detailed descriptions of the Munda languages from a typological perspective are often lacking.

Kharia is often spoken in multi-lingual communities, where its speakers are in daily contact with speakers of Sadri,² the traditional *lingua*

² Or “Sadani”, as it is referred to in most western studies. As the term “Sadri” is preferred in the region, and as “Sadani” is actually a cover term for a number of closely related linguistic varieties, this particular variety will be referred to as Sadri in this study.

franca of the region, and Hindi (both Indo-Aryan), Mundari (North-Munda) and Kurukh³ (North Dravidian). In Orissa speakers of Kharia are also in close contact with speakers of Oriya (Indo-Aryan). All speakers of Kharia whom this author has met are multilingual and speak Kharia, Sadri and Hindi fluently. Occasionally, Kharia speakers will also have some fluency in other languages of the region, such as Mundari, if they have close contact with these other groups. Conversely, speakers of other languages in the region occasionally have some degree of fluency in Kharia if they live in predominantly Kharia-speaking areas. Many Kharia also have some degree of fluency in English.

The standard name of the language and people is *khariya* [kʰə.ɾi.ja], written <Kharia> in English, but many speakers feel that the form *kheɾiya* [kʰɛ.ɾi.ja] is more correct. The etymology of this term is unclear. Konow (Grierson, 1906 [1994]: 190) notes a suggestion which connects the term to the Proto-Munda word for ‘person’. Although this is uncontroversial in the case of the North Munda languages and peoples, who refer to themselves and their languages as either *hoɾ* or *ho* (both of which mean ‘man, person’ in the respective languages), the cognate form in Kharia is *kaɾ*, and there remain a number of problems with such a derivation, especially the aspiration in the name *khariya/kheɾiya*. A number of other popular or “folk” etymologies have also been proposed (see Kullū, 2000: 3–7 for an overview), perhaps the two most interesting of which are the following: In one suggestion, the name is claimed to derive from the Kharia word *khori* ‘village section’, as the Kharia-speaking population often lives together in individual sections of multi-lingual villages. Another derives the name from an Indo-Aryan word for ‘white clay, chalk’, cf. Hindi *khariyā*. The etymology of this term awaits further study.

There are three groups which are generally classified together as Kharia—the *dudh* Kharia (cf. Hindi *dūdh*, Sadri *dudh* ‘milk’), the *delki* or *qhelki* Kharia, and the Hill Kharia. Of these three groups, apparently only the Dudh and D(h)elki Kharia speak Kharia, whereas the Hill Kharia speak Indo-Aryan languages closely related to Bengali and Oriya.⁴

³ Kurukh is also often referred to as “Oraon”. The language is also spoken in some areas in southern Nepal, where it is known as “Dhangar”.

⁴ Malhotra (1982: 7) mentions a group known as Sabbars, in North Orissa, who speak Kharia, although she gives no further data. It is possible that this is in fact a group of Hill Kharia who still speak Kharia, as there is some confusion with respect to the name “Sabbar” and whether or not this is a group of Kharia (cf. e.g. Roy & Roy, 1937: 30f.). Interestingly, Roy & Roy (1937: 31) note that “The Hill Khāriās [sic!], like the present Savara tribe of Ōriṣṣā, have at the present day no language of their own but have adopted the

It is generally assumed that there are two main dialects of Kharia corresponding to the division between the Dudh and D(h)elki Kharia, but this has yet to be confirmed. With the exception of one speaker from northern Orissa, all other speakers of Kharia with whom this author has worked are from the Gumla and Simdega districts of southwestern Jharkhand, all of whom are speakers of the “Dudh” dialect. The present study therefore deals only with Dudh Kharia as it is spoken in southwest Jharkhand and northwest Orissa.

According to the 1971 census, 191,421 people spoke Kharia as their native language at that time, out of a total Kharia population of 321,190 (data from Abbi, 1993: 543), whereas Grimes (1988: 471), quoting another work, gives 111,000–160,000 speakers for roughly the same period. The present Kharia-speaking population has undoubtedly increased since then, although it is not clear how many speakers there are. The internet versions of the Ethnologue⁵ give a total of 278,500 speakers for 1994 (14th edition, 2000) and 292,000 in India for 1997 (15th edition, 2005).⁶

The question of whether Kharia is endangered is difficult to answer. In the short term, it cannot be considered endangered, although the long-term chances of survival are more difficult to assess. As early as Grierson (1906 [1994]: 190) Kharia was declared to be a dying language, a view which also found support in Pinnow (1965a: ix, 4ff.). Nevertheless, the language probably has more speakers now than at any other point in its history due to the size of the ethnic Kharia population, although it seems that an ever-increasing percentage of this population is choosing to rear their children in either Hindi or Sadri, so that many ethnic Kharia have neither an active nor a passive command of the language.

language of their neighbours—namely, *Ōṛiyā* in Mayurbhañj and other Ōṛissā States, and *Bengali* in Singhbhūm (Dhalbhūm), Bankurā and Manbhūm.” For a more recent account of the Savar (*sāvar*, in Hindi), see Bhagat (2001a), who also notes (Bhagat, 2001a: 18) that they may be connected in some way to the Hill Kharia.

Note also that Risley (1891 [1998], II: 241ff.) writes of a tribe named “Savar, Sabar, Saur, Sar, Sayar, Suir, Swiri” and considers them a Dravidian tribe, noting however that other authors consider them “Kolarian”, i.e. Munda, a view which he rejects as he considers the linguistic evidence to be “meagre and inconclusive” (Risley, 1891 [1998]: 242).

The confusion in the literature undoubtedly derives from the fact that there are two languages, Sora / Savara (Munda) and Savar (Dravidian), both spoken in Andhra Pradesh (cf. Lewis, 2009, under “Sora” and “Savara”, respectively), in addition to the presumably Kharia Sabbar of Orissa. Further research on this topic is necessary.

⁵ See note 1 above.

⁶ Note also that the number of speakers given in the 15th edition of the Ethnologue for all speakers in both India and Nepal is only slightly higher than the number of speakers in India: 293,575. The number of speakers for India is given as the same in the 16th edition (Lewis, 2009), although the figure for all countries is slightly higher: 293,580.

There is a growing movement to promote Kharia, both through various organizations as well as in the schools. On the other hand, the increasingly high level of education is bringing the Kharia in their daily lives more and more into contact with Hindi and other languages, while opportunities for speaking Kharia are becoming fewer. Only time will tell to what extent present attempts at preserving the language are successful. Although the language has certainly taken on a large number of traits from Indo-Aryan (cf. Malhotra, 1982 and Abbi, 1993; 1997), it has survived up to the present and the number of people speaking it has even increased substantially in the past 30 years, despite all earlier forecasts.

1.2 GENETIC AFFILIATIONS

The Munda languages together form the western branch of the Austro-Asiatic phylum, which extends from Central India to Vietnam. The phylum, whose genetic relationship was demonstrated beyond any doubt in Pinnow's (1959) excellent overview of the family,⁷ can be divided into two major groups, Munda and Mon-Khmer.

The Munda languages stretch from central India to eastern central India, with a few pockets further to the northeast and elsewhere, due mainly to migration which began in the 19th century. The Mon-Khmer languages stretch from Khasi-War in northeastern India in the northwest, to Vietnam in the east, Malaysia in the southeast and the Nicobar Islands in the southwest.

In the genetic scheme of Munda given in Zide (1969), Kharia and Juang make up the (South-) Central branch of Munda.⁸ This scheme, which has probably been the most influential genetic model of the Munda languages, is illustrated in Diagram 1.1. This model has been questioned by a number of authors, however. Although we will not deal here in any detail with the genetic affiliations of the various Munda languages, it should be noted that in Anderson's (2001) scheme, Proto-Juang and Proto-Kharia(-Gutob-Remo) went separate ways at an early date, although Anderson does note that Kharia and Juang share many common traits. As Anderson (2001: 32) himself notes, the question of internal relationships within South Munda is still very much an unsettled issue.

⁷ In addition to other studies of his on this topic, e.g. Pinnow (1960; 1963).

⁸ For a brief description of Juang, see Matson (1964) and Patnaik (2008).

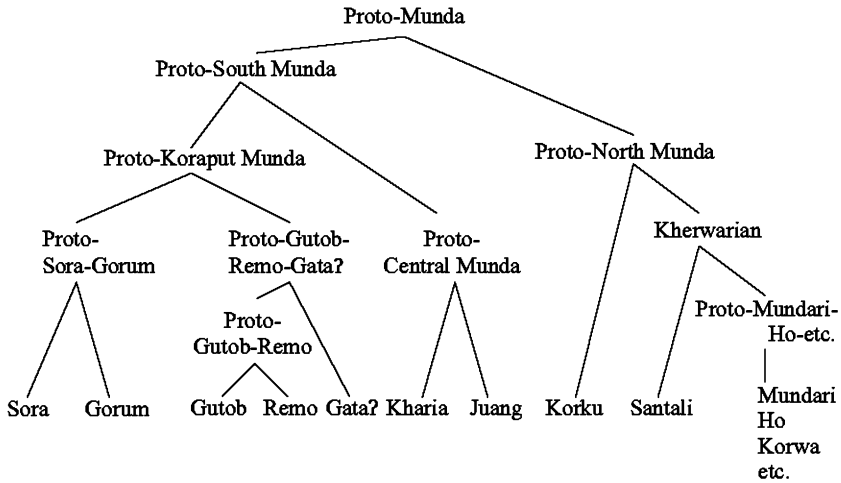


Diagram 1.1: A genetic scheme of the Munda languages (Zide, 1969: 412)

1.3 PREVIOUS STUDIES ON THE KHARIA AND WHAT THIS STUDY ADDS

1.3.1 *Previous works*

The Kharia are one of the best studied Munda groups and studies on their language and culture date back to the 19th century. However, the first major work dealing with the Kharia language is Pinnow (1959), a German-language work on Kharia phonology, predominantly from a historical perspective. Despite the fact that it is now somewhat dated, it is nevertheless still standard reading for anyone interested in historical Munda and Austro-Asiatic linguistics.

A number of other works by Pinnow should also be mentioned here, as they are also basically compulsory reading for anyone interested in Kharia or Munda linguistics: Pinnow (1965a) is a large collection of texts in Kharia, accompanied by a word-for-word translation, a free translation in German and detailed notes. Pinnow (1965b) is a shorter collection of the same type, also in German, but without the word-for-word translation. These last two works provide a valuable resource for anyone wishing to study almost any aspect of Kharia, including to a certain extent language change, as these texts were collected ca. fifty years ago.

Unfortunately, these two works have never received the attention which they deserve: Despite the high quality of the texts, the translations, and the copious notes, for those who do not read German these texts are vir-

tually inaccessible. But even for those who do know German, these texts will be difficult to interpret if the reader does not already know Kharia or is not at least familiar with one other South Munda language. For example, in Pinnow (1965a) each line is accompanied by a word-for-word German translation but unfortunately, this is not a gloss in the usual sense of the term. Rather it is a German translation of the respective (segmented) word, e.g. *po?da-?* ‘des Dorfes’ (English: ‘of the village’) and the reader who is not familiar with Kharia grammar will often have to guess as to the proper analysis. Although in this particular case the difficulties are perhaps minimal (*po?da* ‘village’, *-?* ‘GENITIVE’), in many cases the non-initiated reader will face the almost insurmountable task of determining which part of the German translation refers to which part of the Kharia word(s). Pinnow (1965b), on the other hand, although of equally high quality and carefully commented, contains no word-for-word translation of the text but only the colloquial German translation and notes. This will clearly hinder all who do not already know both Kharia and German from gaining access to these texts.

Roy & Roy (1937), an ethnological study, is a further standard work on the Kharia. It is based on twelve years of research among all three Kharia groups (Dudh, Dhelki and Hill Kharia) by S.C. Roy, often referred to as the “father of Indian ethnography”, and his son. This work (two volumes) contains information on virtually all aspects of Kharia life in the early twentieth century, with the exception of the language. It stands alone in its scope in Munda ethnography and is so detailed that it has even led one modern scholar to write of it that “Later ethnographic accounts have nothing significant to add” (Pfeffer, 1993: 222). Other, more recent ethnological studies in English include Pfeffer (1993) on Kharia totemism, Sinha (1984) on the Hill Kharia, and Vidyarathi & Upadhyay (1980) on all three Kharia groups.

A relatively large number of linguistic studies have also appeared on Kharia, most of which are now primarily of interest from a historical perspective, such as Banerjee (1894 [1982]), the first attempt to describe the language systematically, Tea Districts Labour Association (1929), a short but remarkably precise description of the language, and Floor et al. (1934), also from the Tea Districts Labour Association, a compact English-Kharia, Kharia-English dictionary. While none of these is up-to-date, they have played an important role in the history of Kharia linguistics.

Two further studies on Kharia deserve special mention here: Biligiri (1965) and Malhotra (1982). Biligiri (1965) is the first full-length study of the language, including a Kharia-English lexicon, while Malhotra (1982),

the author's unpublished Ph.D. dissertation, is a comprehensive grammar of Kharia, dealing extensively with all aspects of the language, except phonology, and including a chapter on language change. These two studies, although now also somewhat dated, represent the first detailed descriptions of the structure of Kharia using the methodology of modern linguistics, and the present study owes much to their insights (see the next section for details of the relation of the present study to these two studies).

There are also a number of more recent works on Kharia in English and German, such as Abbi (1993; 1997), dealing primarily with language contact and change, Mahapatra (1976) on Juang and Kharia "verbs", Rehberg (2003), a German-language *Magister*-thesis on Kharia phonology, and Zide & Stampe (1968) on the place of Kharia-Juang within Munda. A number of works by the present author have also recently appeared or are now in press: Peterson (2002a) on finiteness in Kharia; (2005, 2007, in press, a) on parts of speech in Kharia; (2006), this author's *Habilitation*-thesis or "professorial dissertation", consisting of three volumes (altogether over 850 pages)—a grammatical analysis (Volume I), a Kharia-English lexicon (Volume II), and a collection of Kharia texts (Volume III), the first volume being a thoroughly revised and expanded version of an earlier shorter grammar, (2008), which appeared in the multi-contributor work *The Munda Languages* (Anderson (ed.), 2008); (2009), an online Kharia-English dictionary; (in press, b) on the totality marker ("v2") *may*; (forthcoming, a), discussing Kharia's place within the "Jharkhand *Sprachbund*"; (forthcoming, b), a prototypical study of the middle voice in Kharia from a typological perspective, and Peterson & Maas (2009), dealing with reduplication in Kharia.

Finally, although not dealing primarily with Kharia, Anderson (2007) should also be mentioned in this respect, which deals with predicating structures throughout Munda, including Kharia, primarily from a typological perspective. As such, it provides a number of valuable insights into the structure of the TAM/PERSON-syntagma (\approx "verb") in Kharia and Munda languages in general. As the present study is primarily descriptive in nature and is of course primarily restricted to Kharia, we cannot deal with this study further in the present volume, although those interested in Kharia and Munda from a larger typological perspective will find much of interest to them there.

The number of works in Hindi on the Kharia and their language is also significant: Ba? (1983; 2001) deal with Kharia phonology; Bhagat (2001b) is a brief, six-page overview of the Hill Kharia; Đuŋđuŋ (1986) contains

a brief grammar, a large number of Kharia stories and songs, all accompanied by a Hindi translation, as well as a Hindi-Kharia and Kharia-Hindi glossary; Duṇḍuṇ (1999), based on that author's own interviews with the Dudh and Dhelki Kharia, contains a large amount of information on all aspects of Kharia life; Sāhu (1979/80) and Kullū (1981) are concise grammars of Kharia, the latter also with a glossary, and finally Kullū (2000) deals extensively with Kharia religion and culture. Most of these authors who have written in Hindi are themselves Kharia, so that it is also possible now to speak of an indigenous scholarly tradition among the Kharia. In fact, Kharia (culture, language, literature and history) can even be studied at the Department of Tribal and Regional Languages at Ranchi University, where many of the courses are taught by Kharia.

There is also a growing amount of literature in Kharia, beginning with Archer (1942), a collection of songs and riddles in Kharia, although unfortunately only some of these have been translated (see the indices in Pinnow, 1965a: 284, 286). There are also a number of partial Bible translations, song-books for the church service and other religious texts, as well as a translation of some of Premchand's works. Original work in Kharia is also on the increase, ranging from poetry and short stories to drama (e.g. Kerkeṭṭā, 1990, from which many examples in this work were taken), and most recently a novel, the first in Kharia (Kullū, 2007).⁹

1.3.2 *The present study*

The present study is the result of research conducted by the author during five research trips to the state of Jharkhand over the course of nine years, amounting to approximately eight months altogether.¹⁰ It benefits from being able to draw upon earlier works on Kharia, most notably Biligiri (1965) and Malhotra (1982), as well as the extensive corpus of texts in Pinnow (1965a, b), which appeared too early for Biligiri's study, although Malhotra (1982) was able to make use of them. However, the present study also goes beyond the treatment of the language presented in these two studies in a number of respects. In addition to a number of minor issues concerning, e.g., compounds (cf. Section 4.6), animacy as a grammatical category (5.1), the anaphoric proform *adi* (5.6), the "√2s" or

⁹ For a more complete list, see: <http://www.SouthAsiaBibliography.de/Bibliography/Austroasiatic/Munda/Kharia/kharia.html>

¹⁰ I wish to thank the German Research Foundation (*Deutsche Forschungsgemeinschaft*, DFG) for two generous grants (PE 872/1–1, 2) which greatly aided me in my research.

markers of *Aktionsart* (6.5), finiteness (6.6), the “copula” (7.3), “relative clauses” (7.6) and many other more-or-less minor points, the major differences between the present study and these two earlier studies are the following:

- Parts of speech—The major difference between the present study and these two earlier studies lies in its approach to parts of speech. Both of these earlier studies recognize (in addition to a number of other classes) verbs, nouns, and pronouns, while Malhotra also assumes an adjectival class. On the other hand, the present study assumes that there is one OPEN class in the lexicon, consisting of all contentive morphemes, which may all be used in referential, predicative and attributive function (as indeed more complex units or “phrases” may as well), and a CLOSED class, which divides into two further sub-classes: FUNCTIONAL MORPHEMES (case marking, TAM marking, markers for person, number and honorific status, etc.), which may never be used in referential, attributive or predicative function, and a small class of PROFORMS AND DEICTICS, which may be used referentially and predicatively but which require the genitive to be used attributively (cf. Chapter 4). As such, the present study does not make use of terminology such as “noun”, “verb”, “adjective”, or “noun phrase” Also, and most importantly, the present study differs in that it shows that virtually all grammatical marking in Kharia is enclitic (cf. Chapter 3), and not suffixal, and that this marking attaches to complex syntactic units or “(non-endocentric) phrases”, hence what is at issue in Kharia are less LEXICAL CLASSES such as “noun” and “verb” than SYNTACTIC STRUCTURES, termed TAM/PERSON- and Case-syntagmas in this study.
- Basic voice: the active and the middle—The second major difference between the present study and these two earlier studies lies in the area of what Klaiman (1991) refers to as “basic voice”, i.e., active and middle. Biligiri (1965) refers to these two classes as “set 1” and “set 2” suffixes which take “active” and “neutral” bases, respectively, while for Malhotra (1982) these two classes are basically “transitive” and “intransitive” Section 6.4.2 in the present study is an in-depth investigation of these two categories in Kharia which goes into considerably more detail than either of these two previous studies. In addition, this study benefits from recent developments in the field of basic voice and related topics which have appeared since Malhotra’s study.

- Scope—Finally, without wishing to belittle the important contribution made by the authors of these two previous studies, the present volume contains considerably more data on the grammar of Kharia, especially on morphology and syntax, than these two previous studies. Furthermore, as an uncorrected, unpublished manuscript, Malhotra's study can be quite difficult to obtain and contains a rather large number of misprints and other errors, especially with respect to the presence or absence of the glottal stop. The present study is thus more exhaustive in its treatment of morphosyntax, however it does not deal with phonotactics in quite as much detail as Biligiri (1965) nor with language contact / change to the extent that Malhotra (1982) does.

1.3.3 *Methodology*

The data discussed in the following chapters were obtained primarily though a) the analysis of narratives, written short stories and songs, and b) interviews. Data from informal conversations also occasionally find their way into this discussion, although to a much less extent. Also, although the author made every attempt to spend as much time in the countryside as possible, where Kharia is spoken by a large percentage of the population, for reasons of personal safety,¹¹ this was only possible to a very small extent, so that most data had to be obtained working with Kharia speakers in Ranchi, the capitol of Jharkhand. However, as there is a large Kharia-speaking community here, this did not greatly hamper his work, although the data produced in such surroundings, where Hindi clearly dominates, may of course differ to some extent from that used in the countryside.

¹¹ Throughout much of the state of Jharkhand (and other regions of eastern and eastern central India) there is a militant insurgency in the countryside, referred to by various names (Maoists, MCC, Naxalites, etc.). In Jharkhand, these groups are collectively referred to as the MCC (Maoist Communist, Centre), whose official goal is the installation of a Maoist state. During this author's first two trips to Jharkhand, the center of activities of these groups shifted from eastern Jharkhand to western Jharkhand, where Kharia is spoken, gradually making it too dangerous to conduct research in the countryside or to be in the countryside after dark, as kidnappings have become quite common. For this reason, as of the author's third visit, almost all work on Kharia had to be conducted in Ranchi due to its relative safety.

a. *Narratives, written short stories and songs*

Most of the data used in this study are from narratives from various speakers on topics ranging from the history of the Kharia to their everyday lives. Unfortunately, speakers were often reluctant to allow their narratives to be taped without time to prepare them at least mentally (usually one day). There seem to have been many reasons for this, first and foremost that most of the Kharia speakers who this author worked with were also students of Kharia, who were interested in speaking as “pure” Kharia as possible, e.g., using the now virtually defunct Kharia numerals and other words no longer in general use. Also, they often wanted to be sure of details on the history of the Kharia, as these stories are no longer handed down from one generation to the next, so that many students were only familiar with them through their studies at Ranchi University.

There are also a number of written texts in the corpus, which with one exception were all written especially for this author: Although none of the speakers had been asked to write stories, at one point a speaker approached this author with a children’s story which he had composed and which he thought might be of interest. When he found my reaction to be positive, he began writing more stories, and word of this spread to other speakers I had been working with, with the result that I soon had a whole collection of stories to type, gloss and translate, with offers for more. With the exception of the stories written by the person who first presented a story to me (Mr. Basil Baa, who plans to publish these stories), all of these texts were composed by students of the Department of Tribal and Regional Languages at Ranchi University who are studying or have completed their studies in Kharia. With one exception, the texts which were written especially for me are perhaps best considered “fairy tales” or children’s stories, in which bread grows on trees and animals speak and behave as humans. The remaining text, on the other hand, deals with the case of a young man with a drinking problem and his contentious wife, and the problems this causes.

The one written text which had not been written especially for this author is very different from these other texts. It is an account of the mythological wanderings of the Kharia throughout the Middle East and Rome [sic!], and finally through India to their present home in eastern-central India. The story had been handed down orally in the speaker’s family but had been put down in writing at some time.

A number of songs were also included in the corpus. In fact, once I had asked for a few songs, I was virtually inundated with a large selection to choose from, two of which were also composed by the persons who sang

them for me. The corpus contains altogether 30 songs of two types, both the traditional and very common *alon*, in which one person sings a verse which is then echoed by others, as well as the *durāṅ*, which is more ceremonial and which is sung either by one person or by a group together.¹²

Finally, a number of examples from the dramas contained in Kerkeṭṭā (1990) are included here, as these dramas deal with life in the rural areas of the Kharia heartland and hence provide valuable insights into a register of Kharia which could not be obtained in narratives, interviews and conversations.

b. *Interviews*

Several hundred hours of interviews were conducted during the course of the five field trips to Jharkhand, primarily with speakers between the ages of 20 and 35, for practical reasons, although both younger and older speakers were also consulted whenever possible. The language consultants were both male and female, and most were also students or previous students of Kharia at the Department of Tribal and Regional Languages at Ranchi University. Furthermore, for the most part, speakers had come to Ranchi only a few weeks before beginning work with this author from either the district of Gumla or the district of Simdega in southwestern Jharkhand. All were “Dudh” Kharia.

Interviews lasted for ca. 45 minutes and were generally conducted alone with male consultants but generally in groups of two female consultants, who often initially felt uneasy working alone with a foreign man. During my first two visits to Jharkhand in 2001 and 2002, Hindi was the language of consultation, except with Mr. Anil Kullu, with whom English was used. However, beginning with the third trip to Jharkhand in 2003, all interviews were conducted monolingually in Kharia.

As the group of consultants who I worked with on a regular basis became smaller over the course of the years, and as the discussions between myself and this group of speakers became more routine-like, interviews were also often conducted in groups of up to six speakers, especially when dealing with topics such as parts of speech or the meaning of contentive morphemes in the active and middle, topics on which there were often considerable differences of opinion which, as it turned out, were generally due to geographic factors, with speakers from

¹² The entire corpus can be downloaded under: <http://www.southasiabibliography.de/Bibliography/Austroasiatic/Munda/Kharia/kharia.html>

southern and western Simdega district in one group and those from other regions in the other. Where such differences of opinion were noted, the variety of the “Simdega” group is discussed in this grammar in detail, and the other variety, referred to here for ease of reference as the “Gumla” variety, is mentioned in passing.

My main consultants for Simdega Kharia, and with that, for the present grammar, were Mr. Basil Baa, Mrs. Rayem Olem Baa (formerly Dungdung), and Mr. Anil Kullu, and the present grammar owes considerable debt to the unfailing intuitions of these three speakers. Other speakers with whom I also worked extensively include Ms. Tarkeleng Kullu, Mr. Anugrah Kullu, Ms. Rose Dungdung, Ms. Saroj Kerketta, Mr. Silvester Kerketta, Ms. Anita Soreng, Mr. Dilip Soreng, Mr. Pratap Soreng, and Mr. Marianus Tete. All of these speakers were between 20 and 35 during the eight years during which most interviews were conducted. Furthermore, a number of narratives were collected from Mr. Marcus Soreng from Birmitrapur, Orissa, who at the time we met was around 70 years of age. These were then analyzed with the above-mentioned speakers.

I also had a few opportunities to work with a group of young men, aged 17–18, referred to collectively as the “Kharia Boys” of St. Alois Minor Seminary in Ranchi: Sanjay Dung Dung, Edwin R. Dung Dung, Ajit Kerketta, Alexzendar Kerketta, Pavan Prakash Kerketta, Ravi Kujur, Ranjit Kullu, Samir Kullu, Vinay Kullu, Alok Soreng, Anil Soreng, and Shilanan Tete. Through these speakers I gained some insight into Kharia as younger speakers—at the time still in school—speak Kharia.

I also greatly profited from conversations on Kharia, both grammar and lexicon, with the late Mr. Stanislas Kullu, father of Mr. Anil Kullu, and at that time president of the All-India Kharia Society, who also worked with Heinz-Jürgen Pinnow. Mr. Kullu welcomed me into his family and his home in Saldega (*saldə?*) in the vicinity of Simdega for as long as was possible, during which time I was in a largely Kharia-speaking environment and was able to record a number of traditional songs and learn a number of facts about Kharia as it is spoken in the countryside, before having to return to Ranchi for reasons of safety.

1.4 SOME TERMINOLOGICAL ISSUES

There are a number of terminological issues which arise in a description of Kharia which do not arise for many other languages. Although the

terms used throughout this study may seem somewhat irritating to the reader at first sight, they are necessary for a number of reasons which will be dealt with in the following chapters. The reader should have little trouble in adjusting to them quickly, as they introduce few concepts with which most linguists are not already familiar, although their use in a grammatical description is perhaps somewhat unusual. The present chapter introduces these terms only very briefly, serving as a general point of reference for the reader. For further discussion, the reader should consult the relevant chapters mentioned in the text below.

1.4.1 *“Contentive morphemes and semantic bases” vs. “roots and stems”*

As will be shown in chapters 2 and 3, virtually all functional or grammatical markers in Kharia are enclitic, not affixal, although there are a few derivational affixes. These enclitic markers include markers for case, TAM, person/number/honorific status, etc. As such, units which correspond to “roots” and “stems” in other languages often have a different status in Kharia: The supposed “root” in Kharia is a syntactically relevant unit, to which such enclitic grammatical markers may attach in the SYNTAX, hence the term “root” does not seem appropriate, as it implies a unit at a level below that of the “word” which is transformed into a word through derivational and inflectional processes.

Similar comments also hold for supposed “stems” in Kharia, derived from the “roots” via any one of a number of different processes (reduplication, infixation, etc.): These too are syntactically relevant “words” to which enclitics attach in the syntax and as such are not “stems” in the usual sense.

To avoid these terminological problems, the current grammar uses the term “(underived) contentive morphemes” to refer to these most basic, underived units (\approx “roots”), and “derived forms” to refer to contentive morphemes which have undergone some form of derivation (reduplication, reciprocal or causative marking, etc.) (\approx “stems”). Furthermore, as both an (underived) contentive morpheme as well as a derived form can serve as the semantic base of a referential, predicative or attributive unit, the terms “semantic base” or “semantic head” will often be used in this study as a cover term for both contentive morphemes and derived forms.

1.4.2 “Case- and TAM/PERSON-syntagmas” vs. “nouns, verbs and adjectives”

It will be argued in Chapter 4 that, while it may not be impossible to speak of lexical parts of speech in Kharia such as nouns, verbs and adjectives, the simplest description of this language makes no reference to these units, as all contentive morphemes—indeed all syntactically relevant “phrases”—can appear in predicative, referential or attributive function. As such, the present grammar makes no use of these categories. It should be remembered here that—at least from the perspective of the structuralist tradition, to which this author feels bound—the universality of lexical parts of speech is an EMPIRICAL QUESTION, not a self-evident fact of life. As such, those who wish to argue that Kharia can be shown to have nouns and verbs, e.g., with rampant conversion at the level of syntax or with zero verbs which are unique among all morphemes in their syntactic distribution, are free to do so, but the present grammar aims only to present as simple an account as possible for the structures encountered and to make as few theoretical assumptions as necessary in the process.

Viewed from this perspective, if the presence of nouns, verbs, adjectives, etc. is not a necessity for an adequate description of Kharia, it will not do to speak here of “nouns”, “verbs” and “adjectives”—and certainly not of “nouns used as verbs” and “verbs used as nouns”—nor of “noun phrases”, “adjective phrases”, etc. In many ways this is highly unfortunate, as it also entails not being able to employ terms which most readers expect to find.

However, the structure of Kharia presents a simple alternative to these categories. As will be argued in Chapter 4, there are two structurally definable, clause-level units in Kharia which will be referred to as the “Case-syntagma”, as case is its only defining structure, and the “TAM/PERSON-syntagma”, as this unit is defined solely by these categories. Otherwise, the semantic base of these two units has the same structural potential.¹³ Although we cannot go into the functions of these two units at this point in any detail, it should be noted that this is a PURELY STRUCTURAL definition and nothing is implied here with respect to the functions of these two units: Although there is a strong TENDENCY for the TAM/PERSON-

¹³ These terms are motivated by the labels “TAM syntagm” and “article syntagm” for Tongan in Broschart (1997), although there are also considerable differences between the structures in the two languages. I prefer the form *syntagma*, but will use an English-style plural form: *syntagmas*.

syntagma to be used in predicative function and the Case-syntagma to be found in referential function, both units can be found in both functions, and both are also quite common in attributive function. Hence, while the present study does not employ notions such as “noun”, “adjective” and “verb”, no problems should arise for the reader in adapting to the terminology in the present study, as the two syntagmas are referred to purely by their structure.¹⁴

1.4.3 *“Marker of qualitative predication” vs. “copula”*

In Kharia, as in perhaps most other languages, there are two major types of predication, based on the structure of the predicate: narrative and qualitative predication, corresponding roughly to verbal and nominal predication in English. This is discussed in some detail in section 6.1. While the remainder of Chapter 6 is then dedicated to the structure of the narrative predicate (or rather, the TAM/PERSON-syntagma), qualitative predication is dealt with in detail in Section 7.3 where among other things it is argued that there is no copula in Kharia in the usual sense of the term but rather, what translates as some form of the copula (“be”) in English and many other languages is best viewed as a marker of qualitative predication. Hence, these forms will consistently be glossed as “QUAL” in this study to refer to “marker of qualitative predication”

1.5 KHARIA ORTHOGRAPHY AND THE SYSTEM OF TRANSLITERATION USED HERE

Kharia has traditionally only rarely been used in its written form: Although most Dudh Kharia are literate, with some estimates running as high as over 90%, since education is generally in English or Hindi medium and since Hindi is spoken by virtually all Kharia, when the need arises to put something down in writing, this has traditionally been done in Hindi.

When Kharia is written, the Devanagari script is almost always used, which is also used for Hindi, Sadri and a large number of other Indo-Aryan languages. There is also a small movement which aims at introducing an orthography based on the Roman alphabet. This movement, however, is currently very limited in terms of the number of its supporters.

¹⁴ Units may of course also be referred to in a particular context according to whether they appear in predicative, referential or attributive function, where appropriate.

Although almost all Kharia would thus probably agree that Devanagari is the proper medium for writing their language, there is as yet no generally accepted standard orthography, neither 1. for the representation of certain sounds, particularly the glottal stop and the pre-glottalized consonants, nor 2. for the question of when to write “words” together or separately. The following presents a brief overview of the script and the issues involved.

1.5.1 *The script*

The Devanagari writing systems in common use for Kharia differ from one another only minimally. All have in common that each segmental phoneme is generally represented by a single symbol, the one which comes closest to it in the Devanagari script. There are only four regular deviations from this phonemic practice:

1. /d/ and /dʰ/, which are generally realized intervocally as [ɽ] and [ɽʰ], respectively, are written as <ɽ> (ॢ) and <ɽʰ> (ॣ) if they are spoken as such.
2. /g/ in the coda, which is realized as [ʔ] and generally followed by an “echo-vowel” in slow speech, is written as a glottal stop in the fashion that the respective writer chooses to represent this symbol (see below). The glottal stop is also often not written at all, as it is commonly omitted in normal or fast speech.
3. Plosives in the coda are frequently realized as pre-glottalized stops (2.2.2). Here the oppositions [±voiced] and [±aspirated] are neutralized as well as the opposition between dental and retroflex. There are only three pre-glottalized stops: <ᵐb> [ᵐbᵐ], <ᵐd> [ᵐdᵐ] and <ᵐj> [ᵐjᵐ]. Although the pre-glottalized consonants are not phonemes but merely allophones of the corresponding, non-pre-glottalized plosives, writers generally prefer to represent these symbols distinctly by one of the means discussed below.
4. Vowel length, which is specified in Devanagari but which is not phonemic in Kharia, is generally written as it is pronounced, although this is seldomly consistently applied, and writers tend to oscillate between long and short vowels for the same words even in shorter texts. Vowel length is indicated in Devanagari for /u/ and /i/: [ə]~[ʌ] is represented by a “short *α*”, whereas [a] is written as a “long *α*”, again although

this distinction is not distinctive in Kharia. Vowel length is not indicated in Devanagari for /o/ and /e/.¹⁵

The various systems in current use only differ from one another in one respect: The means of representing the glottal stop and pre-glottalized consonants, for which there are no symbols in the Devanagari script. There are four strategies that this author is aware of for dealing with these sounds, which while not phonemic (except for the glottal stop, which is marginally phonemic), are nevertheless very prominent features of the spoken language.

1. Glottalization is not indicated at all. Although this is relatively common practice, this strategy does not seem to have not been adopted by any organization striving for a standardized orthography.
2. The glottal stop is symbolized by writing the short vowel twice to indicate the “echo-vowel”, which often follows the glottal stop in slow speech (e.g. <oo> (ओओ), [ɔʔ°] ‘house’). While this strategy is consistently used to indicate glottalized vowels, it cannot be used to indicate the pre-glottalized consonants. In this system, if pre-glottalized consonants are to be indicated, they are symbolized by writing the corresponding plosive followed by its homorganic nasal, joined together by the symbol “◌̃”. E.g., while [aʔ^a] is written as <aa> (अअ), [ʔb^m] is written as <bṃ> (बम). This seems to be the oldest strategy for representing glottalization and is generally found in literature on or in Kharia which was printed before the year 2000. Since then, it would seem that the following strategy has gained ground in published materials.
3. The glottal stop is consistently realized by the *anugrah* (ॐ), if this symbol is available on the keyboard being used, otherwise the *visarg* (:) is used.¹⁶ This strategy has been adopted by the All-India Kharia

¹⁵ The reasons [ə]~[ʌ] is treated as “short *a*” and [ɑ] as “long *a*” lie in the Sanskrit grammatical tradition, in which “short *a*” becomes “long *a*” in certain grammatical operations. Similarly, /o/ and /e/ in Sanskrit are treated as diphthongs, not monophthongs, and are grammatically related to /u/~/ū/ and /i/~/ī/, respectively. Hence <o> and <e> are inherently long in Devanagari, but are used in Kharia to represent all allophones of /e/ and /o/, regardless of length or quality.

¹⁶ Neither the *anugrah* nor the *visarg* has any phonetic value in Hindi or Sadri, except in Sanskrit (Old Indo-Aryan) loan words. In Sanskrit, both symbols fulfill very different functions.

Society and has recently been propagated by this organization as the correct means of writing the glottal stop. As yet, however, it has not gained general acceptance in every-day use, although it seems to be becoming the accepted norm in printed materials (but see point 4 below). This strategy is consistently used both with glottalized vowels and pre-glottalized stops. Thus [aʔ^o] = <aʔ> (अः), [ʔb^m] = <ʔb> (अब).

4. It has been brought to the attention of this author that a fourth system has not only been proposed for representing glottalization but that this has also been accepted by “the government” as the official orthographical symbol for representing glottalization in Kharia, although it has not yet been possible to confirm this. In this system, the apostrophe “ ’ ” from Roman script is used, whether to represent a glottalized vowel or a pre-glottalized stop. Thus <o’> for [oʔ^o] ‘house’, and <’b> for [ʔb^m].

In the present study, Kharia is represented through a transliteration of the written language, using the symbol <ʔ> for the glottal stop in the coda (or preceding the “echo-vowel”), e.g. <oʔ> for [oʔ^o], and representing pre-glottalized consonants with the symbol <ʔ>, e.g. <ʔb> for [ʔb^m]. Although strategy (4) above would have been a much simpler means for representing the glottal stop, at least from a technical point of view, the vast majority of our texts had already been collected before this recent development had been brought to our attention. Also, following Pinnow and other authors, we do not differentiate between [ə]~[ʌ] and [a], which are transliterated here simply as <a>, nor is vowel length indicated, which is non-phonemic.

This same system of transliteration will also be applied here when quoting examples from other authors and will not be commented on further. Similarly, we also occasionally take the liberty of adapting other authors’ glossing to fit the system used in this study, again generally without further comment.

Table 1.1 presents in three columns the various symbols of the system of transliteration used here and its Devanagari equivalents.¹⁷ The order here follows that used in Peterson (2009) and is based primarily on the order of the English alphabet, with some modifications. Note that this is not the order in which Devanagari is presented in dictionaries, which begin with the vowels, followed by the consonants.

¹⁷ The consonants are given here in Devanagari in their form with the so-called “inherent *a*” or “short *a*”.

Table 1.1: The system of transliteration used here and its Devanagari equivalents

Row 1		Row 2		Row 3	
Transliteration	Devanagari	Transliteration	Devanagari	Transliteration	Devanagari
ʔ	ः, ङ, ञ	h	ह	p	प
a	अ, आ	i	इ, ई	ph (= [f], [pf])	फ
b	ब	j (= [j], [dʒ])	ज	t	ड
bh	भ	jh (= [jʰ], [dʒʰ])	झ	tʰ	ढ
c (= [c], [tʃ])	च	k	क	r	र
ch (= [cʰ], [tʃʰ])	छ	kh	ख	s	स
ɖ	ड	l	ल	t	ट
ɖʰ	ढ	m	म	tʰ	ठ
d	द	ŋ	ङ	t	त
dh	ध	ɲ	ञ	th	थ
e	ए	ɳ	ण	u	उ, ऊ
g	ग	n	न	w	व
gh	घ	o	ओ	y (= [j])	य

Aspiration is represented in this system by <h>, not <ʰ>. Where the system used here and IPA differ considerably, the corresponding IPA value is given in parentheses in the table.¹⁸

The Devanagari system is in fact much more complex than this simple table suggests, as the vowels have different representations, depending on whether or not they are preceded by a consonant in a written word. The following presents a few examples to demonstrate this.

1. No preceding consonant in the written word: <a> अ <i> इ <u> उ
 Preceding consonant in the written word: <ta> त <ti> ति <tu> तु

Also, a number of consonants and vowels have different forms when they combine with certain other segments, especially <r>. For details, the reader is referred to the discussion of the writing system in any introductory Hindi course, as the principles involved are essentially the same in Kharia as those used in Hindi orthography.

¹⁸ I use the IPA symbols to represent the retroflex/postalveolars, not the Indological system, which is found in many grammars of South Asian languages. Thus I write <ɖ>, <ɖʰ>, <ʈ>, <ʈʰ>, <ɟ>, <ɟʰ> and <ɳ> instead of <ḍ>, <ḍʰ>, <ṭ>, <ṭʰ>, <ṣ>, <ṣʰ> and <ṇ>, respectively. I also represent the velar nasal as <ŋ> instead of the Indological sign <ṅ> and the palatal nasal as <ɲ> instead of <ṅ̌>.

1.5.2 *The written word in Kharia*

There is also the issue of the written word in Kharia, i.e., when to write two linguistic units separately or together, which to our knowledge has not yet been standardized for Kharia. The following principles will be adhered to here. Details on the status of affixes, enclitics, pre- and postpositions as well as the phonological word are discussed in detail in Chapters 2 and 3.

- Phonological words are written here as separate words. Phonological words are those units which have the typical low-high pitch pattern, including all lexical elements and pre- and postpositions (cf. Section 2.5). Clitics will be considered here to be part of the written word and are adjoined to their hosts here by the sign “=”
- Elements in compounds (4.6) as well as (derivational) affixes (3.3) are adjoined to the preceding or following element by the hyphen “-”
- Demonstratives and interrogatives are generally written as separate words, e.g. *ho o?* ‘that house’, *i o?* ‘what house?’, except when they have proclitic status, in which case they are joined to their hosts by the sign “=”, e.g., *u=ghay* [that=way] ‘thus’, *i=te* [what=OBL] ‘where?’

1.6 THE STRUCTURE OF THIS STUDY

The remainder of this volume is structured as follows: Chapter 2 deals with phonological issues, beginning with vowels (2.1) and consonants (2.2), syllable structure (2.3) and morphophonology (2.4). 2.5 then deals in some detail with the issue of phonological words in Kharia while Section 2.6 briefly deals with sentence prosody.

Chapter 3 begins with a discussion in 3.1 of the distributional characteristics of the morphosyntactic enclitics, arguing that these units not only behave phonologically as enclitics, as shown in 2.5, but also syntactically as well. These are then classified in 3.2 according to various criteria, while 3.3 presents a brief discussion of the few derivational affixes of the language.

Chapter 4 deals with the lexicon, predominantly with the issue of lexical classes or parts of speech. After a brief introduction (4.1), the “pre-categoriality” of both simple contentive morphemes (4.2) and derived forms (4.3) is discussed in detail. 4.4 then shows that the issue is in fact much more complex in Kharia, as what is involved here are not lexical roots or stems but rather syntactic structures. This information is then briefly summarized in 4.5. 4.6 deals with what are often considered com-

pounds in Kharia, where it is argued that most of the structures presumed to be compounds in previous literature are in fact merely juxtaposed contentive morphemes which do not form a tighter unit, although there are a few genuine compounds in the language, at least from a historical perspective. Finally, 4.7 deals with a number of constructions which can be subsumed under the term “echo-words” This term, as it is generally used in South Asian linguistics, refers to what is perhaps more commonly known under the name of “melodic overwriting” in general linguistics. However, this section will go beyond melodic overwriting in the usual sense to include a number of patterns of synonym and antonym repetitions as well as other related patterns, hence the more general heading.

Chapter 5 is a discussion of the structure of the Case-syntagma (\approx “NP”) organized for ease of reference more-or-less along the lines of a discussion of the NP in traditional grammars, dealing with grammatical categories such as case, number, inalienable possession, modification, simple, interrogative and indefinite proforms, quantifiers, etc.

Chapter 6 deals with the structure of the TAM/PERSON-syntagma (\approx “verb”). It presents a discussion of the possible structures of the semantic base of the TAM/PERSON-syntagma and the basic person / number / honorific and TAM categories (6.1–6.4.1). This is followed by a detailed discussion of the active and middle in 6.4.2, two categories which pervade almost the entire TAM-system. 6.5 then provides an exhaustive listing of the so-called v2s, i.e., grammatical markers denoting the passive / reflexive and a variety of *Aktionsart* categories, markers which often derive from contentive morphemes. The remainder of this chapter then deals with partially finite and non-finite forms (6.6), negation (6.7), periphrastic TAM/PERSON-syntagmas (6.8) and inference (6.9).

In Chapter 7, the major syntactic operations and categories are discussed in detail, such as grammatical relations (7.1), interrogatives (7.2), qualitative predication (“nominal” or “copular” sentences) (7.3), clause coordination and subordination (7.4–7.5), a detailed look at the numerous strategies for expressing propositional attribution or “relative clauses” in Kharia (7.6) and closing with a discussion of information structure (7.7).

Finally, a short narrative text on Kharia mythology is presented in the Appendix, dealing with the origin of the nine major Kharia clans.

2.1 VOWEL INVENTORY

2.1.1 *Monophthongs*

Kharia has the following five distinctive monophthongal vowel phonemes, given here along with their most common allophones. It is not (yet) possible to provide exact rules with respect to their distribution, except that long vowels are often found in word-final position in open syllables. There is also a good deal of speaker-specific variation in our data, as we worked with speakers from a number of different regions.

- | | |
|-------------------|-------------------|
| 1. i [ɪ, ɨ, i(:)] | u [ʊ, ʊ, u(:)] |
| e[ɛ, ɛ, e(:)] | o [ɔ(:), ɔ] |
| | a [ɑ(:), a, ə, ʌ] |

Biligiri (1965: 18) and Pinnow (1959: 29) consider nasalization (marginally) phonemic, while it is considered non-phonemic here, as we know of no genuine minimal pairs which are distinguished by nasalization alone. Pre-nasal vowels tend to be nasalized, but not consistently so. In (2) a few minimal and pseudo-minimal pairs are given to demonstrate the status of these vowels as phonemes.

2. /a/ vs. /i/: *anaŋ* ‘we (dual, inclusive)’ / *aniŋ* ‘we (plural, inclusive)’
jayɖam ‘become eager’ / *jayɖim* ‘wake up’
bunam ‘snake’ / *bujim* ‘the person one is named after’

/a/ vs. /o/: *jaʔ* ‘pull out of the ground’ / *joʔ* ‘sweep’
 =*aʔ* ‘genitive-case marker’ / =*oʔ* ‘past active marker’
aloŋ ‘sing; song’ / *oloŋ* ‘straw for thatching rooves’

/a/ vs. /u/: *ulaʔ* ‘leaf’ / *uluʔ* ‘boil’

/e/ vs. /a/: =*te* ‘present active marker’ / =*ta* ‘present middle marker’
ter ‘give’ / *tar* ‘kill’
ɖeʔj ‘chop’ / *ɖaʔj* ‘feel cold (of water, etc.)’

/i/ vs. /e/: *jim* ‘thatch’ / *jem* ‘warm’
dʒɪn ‘draw, pull’ / *dʒɛn* ‘cook (pulse, vegetables)’
dʒil ‘be covered’ / *dʒel* ‘come’

/o/ vs. /e/: *kolon* ‘bread’ / *koleŋ* ‘king’
goʔj ‘die’ / *geʔj* ‘pluck (fruit)’

/o/ vs. /u/: *koyo* ‘wind’ / *kuyu* ‘pot’
moʔ ‘smoke’ / *muʔ* ‘emerge’
roʔ ‘spill’ / *ruʔ* ‘open’
goʔj ‘die’ / *guʔj* ‘puff (of grain)’
poʔd ‘sprout, grow’ / *puʔd* ‘jump, spring’
oson ‘bitter’ / *osun* ‘bewitch’

/i/ vs /o/: *tomsin* ‘wasp, hornet’ / *timson* ‘fire’

/i/ vs. /u/: *gil* ‘beat’ / *gul* ‘cry out’

2.1.2 *Diphthongs*

Rehberg (2003: 6) lists the following five units as diphthongs. The phonetic values given here are their most common realizations. The rules governing the distribution of the various allophones of /ae/ are as yet unclear, other than that [əɪ] appears to be restricted to loan words from Sadri.

3. /ae/ [aɛ, əɛ, əɪ], /ao/ [aɔ], /ou/ [ɔɪ], /oi/ [ɔɛ] and /ui/ [ɯɪ]

As a transliteration of the written language is used here, which is somewhat idiosyncratic in this regard, these will be written as *ay*, *aw*, *ou*, *oy* and *ui*, respectively, unless they occur in closed syllables, where the first two are transliterated as *ai* and *au*.

The status of these units as diphthongs is doubtful: It seems more appropriate to treat them simply as a vowel in the nucleus and a glide in the coda rather than as diphthongs, at least in the native vocabulary. This is primarily due to the incompatibility of these combinations with a consonant in the coda (cf. Rehberg, 2003: 17f.). Although such examples are found, all of these are either loan words from Indo-Aryan (generally Sadri) or place names of uncertain origin: *cair* ‘four’, *j(h)aiit* ‘animal’, *sou²b* ‘all’, *bhoir* ‘entire’, *buidh* ‘intelligence’ (all of Indo-Aryan origin) and *digduin* ‘Digduin (name of a town)’

2.2 CONSONANTS

Table 2.1 presents the consonant phonemes of Kharia, given in what is now more or less the standard transliteration. The status of forms given in parentheses “()” as phonemes is uncertain. Where generally accepted transliteration and IPA values differ, the IPA value is given in brackets “[]” For example, the phoneme /f/ is regularly written in Devanagari as “फ”, which is transliterated as <ph> (1.5.1). Hence, although this grapheme may represent either of the two allophones of /f/, i.e., [f] or [pf], it is transliterated here as <ph>, following general practice. Aspiration is indicated in the transliteration by an “h”, e.g. *th* is pronounced [tʰ], *bh* as [bʰ], etc.

Comments on Table 2.1:

- What are generally referred to as retroflex plosives in studies on Kharia are most often realized as postalveolars. Following general practice, however, they will be represented and referred to here as retroflexes. Their exact status awaits further study.
- [ɽ] is only marginally phonemic. It is normally an intervocalic allophone of /d/: *dɔʔd=na* [take=INF] ‘to take’ vs. *dɔɽ=e* [take=ACT.IRR] ‘take!’ There is, however, at least one minimal pair: *oɔɔʔ* ‘become more; more; and’ vs. *oɽoʔ* ‘get stuck’ [ɽʰ], on the other hand, is restricted to Indo-Aryan loan words and is an intervocalic allophone of /dʰ/.
- /f/ (written as <ph>) is generally realized as [f] but is pronounced by some speakers as the affricate [pf].
- /c(h)/ and /j(h)/ are often realized as the affricates [tʃ^(h)] and [dʒ^(h)], especially in loan words.
- [ʔ] is generally an allophone of /g/: [g] appears in the onset, [ʔ] in the coda. For example, a morpheme-final [ʔ] often alternates with [g]

Table 2.1: The consonants

	Labial	Dental	Retroflex / Postalveolar			Palatal	Velar	Glottal
Plosives	p b bh	t d dh	ʈ ɖ	tʰ dʰ	c ch j [j] jh [jʰ]	k kh g gh	(ʔ)	
Nasals	m	n	(ɳ)		ɲ	ŋ		
Flaps		r	[ɽ]	(ɽ)	(ɽh)			
Approximants	w	l			y [j]			
Fricatives	ph [f]	s					h	

when an enclitic with an initial vowel follows, e.g. *o?* ‘house’, *og=a?* [house=GEN] ‘of the house’

However [ʔ] is not always an allophone of /g/. For example, for some speakers [ʔ] can also serve as an infix which marks the causative with polysyllabic contentive morphemes and which usually has the form [ʔb], an allophone of /b/ (see below). For example, *botoŋ* ‘fear’, *bo<ʔ>toŋ* or *bo<ʔb>toŋ* ‘scare’ This suggests that [ʔ] is either a (somewhat marginal) phoneme in its own right or perhaps an allophone of /b/.

The glottal stop is typically followed by an “echo-vowel” in slow, careful speech, i.e., the vowel preceding the glottal stop is repeated as an extra-short vowel after the glottal stop: /gaʔ/ ‘rip’ is then pronounced [gɑ.ʔɑ], which will be represented in the following as [gɑ.ʔɑ] to stress that this “echo-vowel” is not a phonological syllable. Although the “echo-vowel” produces a second syllable, phonetically speaking, this is not a syllable from a phonological point of view and operations which are sensitive to the number of syllables of the contentive morpheme treat forms such as [gɑ.ʔɑ] as a single syllable.¹ In fast speech the “echo-vowel” (and, less often, the glottal stop) is generally omitted.

The following presents minimal pairs for a number of the phonemes presented in Table 2.1.

4. /g/ vs. /k/: *goŋ* ‘cook’ / *koŋ* ‘know’, *gone* ‘tooth’ / *kone* ‘mouse’
 /b/ vs. /bh/: *bu?* ‘beat a drum’ / *bhu?* ‘bark’ (v.)
 /t/ vs. /th/: *tomsin* ‘wasp’ / *thomsin* ‘9’
 /p/ vs. /ph/: (near minimal pair) *pampla* ‘leaves which have fallen to the ground’ *phamphla?* ‘butterfly’
 /d/ vs. /dʔ/: *bida* ‘farewell’ / *biɖa* ‘rice seed’, *kuda* ‘millet’ / *kuɖa* ‘a kind of black berry’
 Near minimal pair: *daru* ‘tree’ / *ɖaru* ‘fool’
 /t/ vs. /dʔ/: *tuta* ‘bottom’ / *tuɖa* ‘tomorrow’
 /ɲ/ vs. /ŋ/: *eŋ* ‘open (of clouds), clear’ / *eŋ* ‘return’ (ITR)
 /m/ vs. /n/: *ɲom* ‘bask in the sun’ / *ɲon* ‘drive (an ox-cart)’
 /l/ vs. /ɫ/: *gul* ‘cry out’ / *gur* ‘fall’

¹ See the discussion of the masdar in Section 6.6.2.1.

- /r/ vs. /s/: *irin* 'plum' / *isin* 'cook (rice)'
 /r/ vs. /dʒ/: *iriʔb* 'forget' / *idʒiʔb* 'night'
 /dʒ/ vs /j/: *goʔd* 'culminatory telic marker' / *goʔj* 'die'
 /t/ vs /s/: *toloʔ* 'threshing floor' / *soloʔ* 'dog'
 /m/ vs /bh/: *kimin* 'daughter-in-law' / *kibhin* 'thick, dense (of forest)'
 /ʔ/ vs /r/: *thuɾay* 'tap s.o. on the shoulder' / *thuray* 'collect, gather';
koɾi 'waist' / *kori* '20'
 Ø vs /ʔ/: *kunɖa* 'pot for rice-beer' / *kunɖaʔ* 'husk'
 /m/ vs. /ŋ/: *ɔnem* 'warm oneself' / *ɔneŋ* 'give back', *ɖam* 'arrive' /
ɖaŋ 'send'

The distinction between /m/ and /ŋ/ is often neutralized in word-final position, *thom* / *thonj* 'for; PURP' and, less seldomly, in intervocalic position as well, *tomon* / *toŋon* 'stand'

2.2.1 Gemination

Geminates of most consonants are allowed but only at morpheme boundaries: /oton=na/ [ɔtɔn:ɑ] [press=INF] 'press' Pairs such as the following show that consonant length, unlike vowel length, can be phonemic:

5. *u=naʔ* 'of this' *un=na=ʔ* 'to place' (GEN)
 this=GEN place=INF=GEN
6. *koleʔj* 'quarrel, fight' *kol* *leʔj* 'curse one another'
 REC curse

There are also various cases in which gemination may be assumed to have been lost, such as *koleʔj* from *kol leʔj* above, and *umay* 'NEG.3PL', from **um=may* 'NEG=3PL'

/ʔ/, /s/ and /h/ may not be geminated.

2.2.2 Pre-glottalized stops

Plosives in the coda are often realized as pre-glottalized stops, especially in slow, careful speech. In this position the oppositions [±voiced] and [±aspirated] are neutralized as well as the opposition between dental and retroflex plosives. There are three pre-glottalized stops: /ʔb/ [ʔbʷ], /ʔd/ [ʔdʷ] and /ʔj/ [ʔjʷ]. There is no /ʔg/: When /g/ occurs in the coda, it is realized as [ʔ], except in recent loan words, where it is pronounced as [g].

The pre-glottalized stops are pronounced as follows: As the flow of air is suppressed for the production of the plosive, the glottis contracts, producing a glottal stop. Before the plosive is (non-audibly) released, the glottal stop is released, and air flows through the nasal cavity, producing a homorganic nasal. Often, however, no nasal component nor glottalization is audible, resulting in the forms [b̠], [d̠] and [j̠], respectively.

In order to differentiate between the (marginally phonemic) /ʔ/ and the (non-phonemic) pre-glottalized stops, the former will be consistently written as <ʔ> while the latter will be represented as <ʔb>, <ʔd> and <ʔj>. For example, what is written <poʔda> ‘village’ has the phonemic structure /poʔda/ or, perhaps, /pogda/. On the other hand, what is written <ɖoʔɖna> ‘to carry’ has the phonemic structure /ɖoɖna/.

Although the glottal stop is not – or at best is only marginally – phonemic, its presence in the text will consistently be indicated, as well as that of the pre-glottalized stops. Although from a linguistic point of view this may seem unnecessary, there is one very strong argument against not indicating it and we feel that this argument is important enough to override all other considerations: Glottalization is highly characteristic of spoken Kharia and when Kharia is written, authors will generally indicate in some manner that a consonant in the coda is pre-glottalized and will virtually always use an unambiguous marker for the glottal stop – <g> is never used to represent [ʔ]² and we do not believe that a transcription which does not take this into account would be accepted by speakers of the language. For this reason, glottalization will consistently be indicated.

2.3 SYLLABLE STRUCTURE

Kharia words of native origin all have the syllable structure (C)V(C), where “C” is a consonant and “V” a vowel or nasal.³ In addition, there are the following general constraints:

- Words may begin with any consonant other than /ɽ/, /ɽh/ or /ŋ/ (Rehberg, 2003: 15). /w/ is only found word-initially in Indo-Aryan loans.

² As a matter of fact, speakers who this author suggested this to were adamantly opposed to writing <g> for [ʔ].

³ Loan words from Indo-Aryan generally retain their syllable structure, resulting in consonant clusters in both the onset and the coda as well as diphthongs in closed syllables.

- [ʔ] occasionally occurs in the onset in the spontaneous speech of some speakers (Rehberg, 2003: 23) but is never phonemic in this position and is in free alternation with an empty onset.
- The oppositions [±voiced], [±aspirated] and dental/retroflex are neutralized in the coda (2.2.2).
- There are no native words with an /s/ or /h/ in the coda. Where these are found, the contentive morpheme is invariably of Indo-Aryan (usually Sadri) origin, such as *bes* ‘good’, *jinis* ‘animal, thing’, *sandeh* ‘doubt’, etc.
- The nucleus may consist of a vowel or nasal. /ŋ/ is attested in this function in our data: *laphŋga* ‘cave’ [lə.pfŋ.gə]. Similarly, *dhohŋge* ‘lean back and relax’ and *lorŋgob*, an alternative pronunciation of *loren* ‘clean, pure’ Rehberg (2003: 14) also cites the form /pʰophŋdʌ/ ‘fungus’ from Biligiri (1965: 11), presumably with [ŋ] in the nucleus.

2.4 MORPHOPHONOLOGY

Morphophonological phenomena can conveniently be divided into two groups.

1. In the first group, these are restricted to a particular morpheme or morphemes.

The past, active marker =oʔ

Before this marker a morpheme-final plosive is devoiced and aspirated, *oʔj* ‘drive’, *och=oʔ* ‘s/he drove’ Morpheme-final [ʔ] (< /g/) is realized as [kʰ]: *joʔ* ‘eat’, *jokh=oʔ* ‘s/he ate’ (6.4.1.1).

After vowels, the glide -y- [j] obligatorily appears before the past active marker to avoid a hiatus and, for some speakers, optionally after consonants as well. The past, active morpheme in these environments is considered here to have the form =yoʔ and is segmented as such: *thisa=yoʔ* ‘s/he called out’, *yo=yoʔ* ‘s/he saw’⁴

The first and second persons in the past active have slightly irregular forms: =oʔj ([ɔʔj^m] < *=oʔ=j) and =oʔb ([ɔʔb^m] < *=oʔ=b), respectively.

⁴ Similarly, where the glide -y- appears before other enclitics beginning with a vowel, such as =e ‘=ACT.IRR’ (see below), it is also considered to belong to the enclitic.

ḍel ‘come’ and *col* ‘go’

The final /l/ in *ḍel* ‘come’ and *col* ‘go’ (< Sadri *cal-* ‘go’) is dropped before all enclitics beginning with /n/, e.g. *co=na* ‘to go’ (infinitive), *ḍe=na=ki* [come=MID.IRR=PL] ‘they will come’ vs. *col=ki* [go=MID.PST] ‘s/he went’ and *ḍel=ki* [come=MID.PST] ‘s/he came’ This does not hold for other contentive morphemes ending in /l/: *ol* ‘bring’, *ol=na* ‘to bring’ (infinitive).

The perfect

As noted already in Biligiri (1965: 59), the perfect marker =*siʔ(d)* can be realized as [chr.ʔ] after a morpheme-final /j/, /goj=siʔ/ ⇒ [ʔgɔʔj̄.cʰɪ.ʔ] ‘s/he has died’ However, this is not obligatory, and the pronunciation [ʔgɔʔj̄.sɪ.ʔ] is also acceptable.

2. The second group is more phonologically conditioned, i.e., the changes here are due only to the phonotactic rules of the language and are not restricted to particular morphemes.

/g/ > [g] / [ʔ]

When a morpheme ending in [ʔ] (< /g/) is followed by an enclitic beginning with a vowel (except for the past active marker =*oʔ*, see above), [ʔ] may be realized as [g], e.g. *oʔ* ‘house’, *og=aʔ* ‘house=GEN’, *noʔ=te* [eat=ACT.PRS] ‘s/he eats’ vs. *nog=e* [eat=ACT.IRR] ‘s/he will eat’ However, this is not obligatory for the genitive, and most speakers tend to insert the glide -y- [j] after the glottal stop (and other plosives) and before the genitive marker =*aʔ*, e.g. *ḍaʔ-y-aʔ* ‘of the water’ This process has spread to most environments, so that it is now probably more appropriate to consider the underlying form of the genitive marker to be =*yaʔ* instead of the etymologically expected form =*aʔ*.

No complex onsets or codas

The progressive markers =*teʔjḍ* (active voice) and =*taʔjḍ* (middle voice) (6.4.1.3) are realized as such only before the markers of the first and second persons, singular, the only two personal markers which begin with a vowel. Otherwise the final *ḍ* is dropped: *likha=teʔjḍ=ijn* ‘I am writing’ vs. *likha=teʔj=ki* ‘they are writing’ This is due to the strict (C)V(C) structure of the native elements of the language, i.e., complex onsets and codas are not allowed. The same distribution holds for the present-tense qualitative predicative markers (≈ copulas) *aviʔjḍ(d)* (non-negative) and *umboʔiʔjḍ(d)* (negative) (7.3).

- A rather large but limited number of elements, the phonological and syntactic clitics, do not show this pattern. They may have a falling or rising pitch in a particular utterance, but they do not have the obligatory LH pitch pattern of the elements just mentioned.

In our analysis, a phonological word in Kharia minimally consists of one unit with an LH-pitch pattern and often one or more functional elements, the enclitics.⁷ We can therefore define the phonological word in Kharia as that unit beginning with an LH-pattern which continues until either the next unit showing this LH-pattern or until a pause (cf. Peterson & Maas, 2009: 221). Diagrams 2.1–2.3 illustrate this basic LH-pattern.⁸ In addition, Diagram 2.1 also shows that the pitch may drop off in the last syllable of the contentive morpheme, which is quite common when words are pronounced in isolation, as in this case, but only after having peaked.

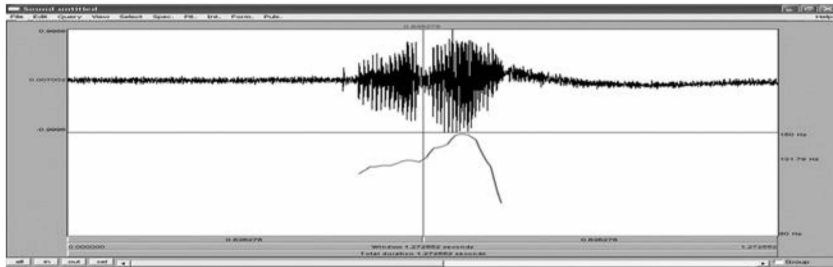
There is no expiratory accent in Kharia and any syllable of the contentive morpheme may be more “prominent” with respect to intensity (cf. Rehberg, 2003: 24ff. and 53ff. (tables)). Although the intensity does tend to be highest in bisyllabic contentive morphemes on the last syllable, where the high tone is found, it is quite often the case that the first syllable is pronounced with a higher intensity than the last syllable, and often the two syllables are approximately equal in terms of intensity. On the other hand, the pitch pattern described above is ALWAYS found in the contentive morpheme.

As Diagrams 2.4 and 2.5 show, even when the tone begins somewhat higher at the beginning of a contentive morpheme, it immediately drops down before rising again on the second syllable. This is especially typical of words beginning with /s/ as well as words beginning with some vowels, such as /o/.

⁷ In Rehberg’s (2003: 23ff.) analysis, Kharia has a non-distinctive pitch accent, marked by a low-tone pitch on the first syllable of the word. After this syllable, the pitch then rises on the following syllable. Thus, in her analysis, in the word *rochoʔb* ‘side’ [rɔ.ɕʰɔʔb^m] the first syllable would carry the low-tone pitch accent, and the second syllable would then be unaccented. In contrast, I do not assume a low-tone pitch accent for Kharia, among other reasons due to the fact that monosyllabic words, such as *lanj* in Diagram 2.3 with a rising pitch but no low tone, would then have no accent. Rather, I propose that the LH pitch pattern identified by Rehberg defines the beginning of a prosodic word, usually the semantic base.

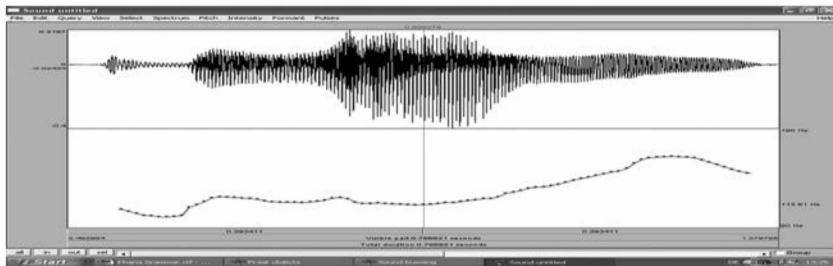
⁸ To keep the following diagrams simple, intensity is not indicated in the lower half of these diagrams, only pitch. See the comments in the following text on intensity. Diagrams 2.1, 2.2, 2.3, 2.7 and 2.12 have been taken from Peterson & Maas (2009: 222). With kind permission of Springer Sciences and Business Media, which is gratefully acknowledged here.

Unlike *kol* 'REC' (6.3.5) and contentive morphemes, which all have an inherent LH-pattern, the enclitics, which denote categories such as inalienable possession, number, case, TAM/basic voice or PERS/NUM/HON-marking,⁹ have no inherent pitch pattern. Cf. Diagrams 2.6 and 2.7.



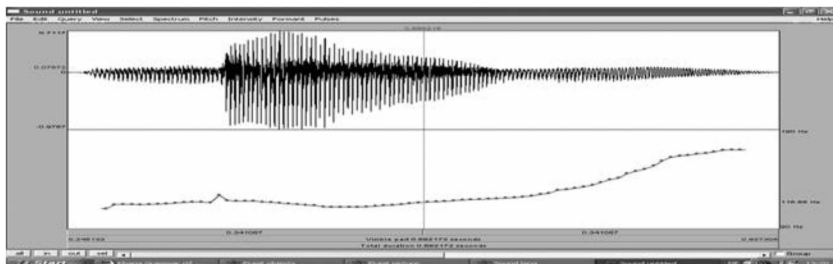
o r e ʔ j

Diagram 2.1: *oreʔj* 'ox'



b u w a ŋ

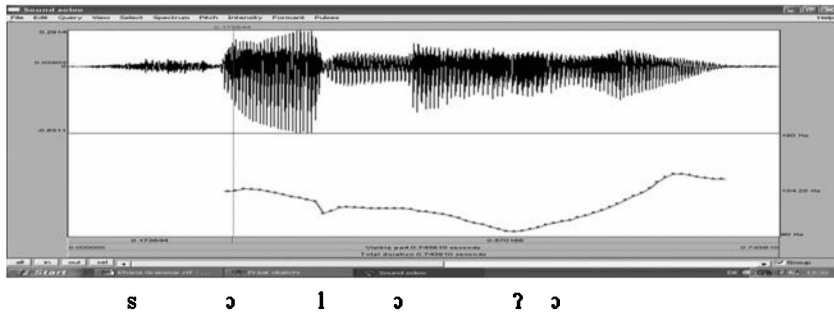
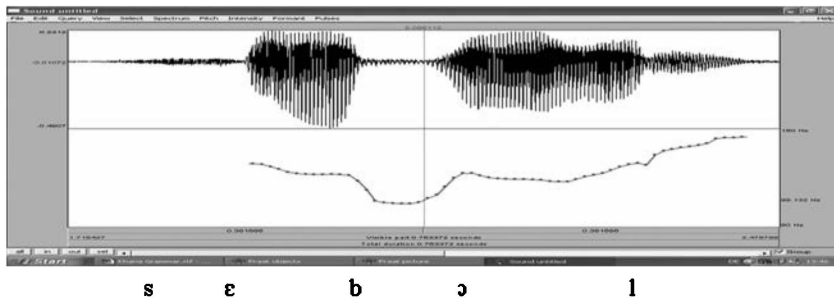
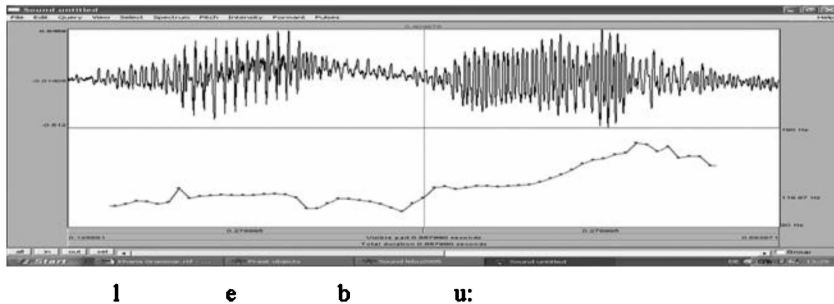
Diagram 2.2: *buwaŋ* 'snake'



l a: ŋ

Diagram 2.3: *laŋ* 'tongue; language'

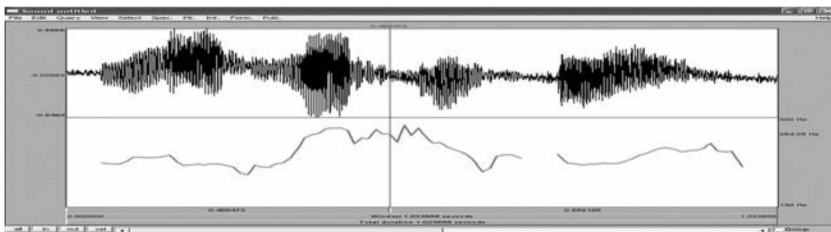
⁹ Cf. Chapter 3.

Diagram 2.4: *solo?* 'dog'Diagram 2.5: *sebol* 'sweet, tasty'Diagram 2.6: *lebu* 'man; person'

In Diagram 2.6 we find the usual LH-pattern of the contentive morpheme. In Diagram 2.7, we again find the LH-pattern with *lebu*. However, number and case markers behave markedly differently: =*ki* has a clear falling pattern here, unlike *lebu*. On the other hand, =*te* rises slightly, due to the

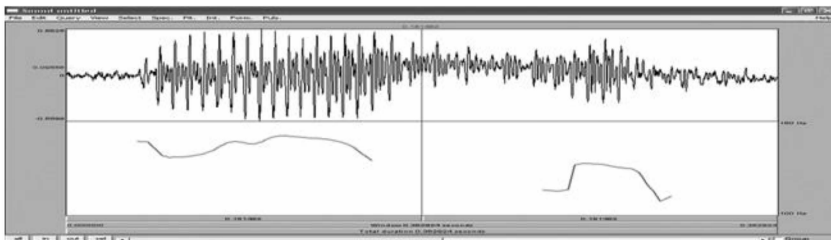
fact that when words are spoken in isolation (as here), the last part of the word is pronounced either with a falling or a rising pattern. However, the contentive morpheme *MUST* be pronounced with a rising pitch.¹⁰

Consider now Diagram 2.8, which shows the pitch pattern in a monosyllabic contentive morpheme. The pitch in *o?* rises, as expected, although it drops before [ʔ]. The “echo-vowel” [ɔ] is pronounced with a lower pitch, which is quite common following the glottal stop but is not of importance here, since the pitch reaches its peak before the glottal stop. Recall also that, phonologically speaking, the “echo-vowel”, here [ɔ], does not count as a second syllable (and is often not pronounced, cf. 2.2).



l e b u ʃ i t e

Diagram 2.7: *lebu=ki=te* [person=PL=OBL] ‘people (OBJECT)’



ɔ

ʔ

Diagram 2.8: *o?* ‘house’

¹⁰ Most examples were read twice by speakers. Generally, the first reading contained a rising pitch at the end of the utterance, whereas the second reading had a falling pitch after the contentive morpheme. Crucially, however, this never affects the pitch pattern in the contentive morpheme itself, which is always LH. Cf. also *=te* in Diagrams 2.9, 2.11 and 2.12.

As Diagrams 2.9-2.12 show, enclitics behave similarly with most monosyllabic morphemes (at least those ending in /ʔ/) as with bisyllabic morphemes.

Often, however, enclitics may combine with a monosyllabic host which does not end in /ʔ/ to form an LH-pattern. Consider Diagram 2.13: Here,

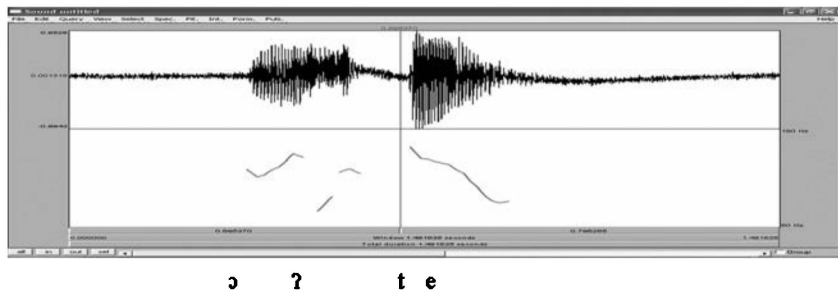


Diagram 2.9: $oʔ=te$ [house=OBL] ‘in the house; the house (OBJECT)’

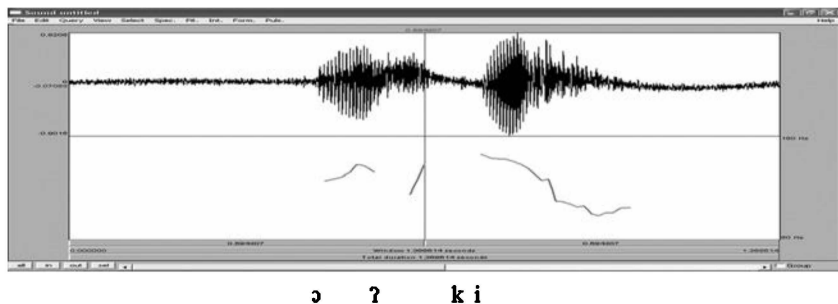


Diagram 2.10: $oʔ=ki$ [house=PL] ‘houses’

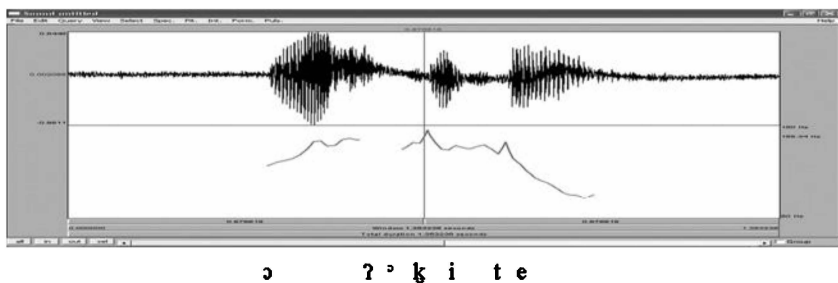


Diagram 2.11: $oʔ=ki=te$ [house=PL=OBL] ‘in the houses; the houses (OBJ)’

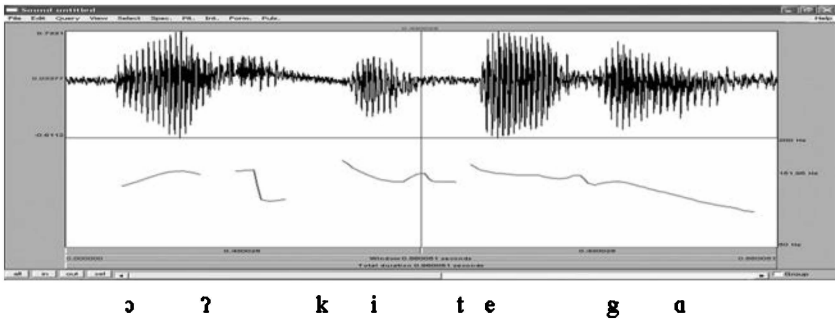


Diagram 2.12: *o? = ki = te = ga* [house=PL=OBL=FOC]

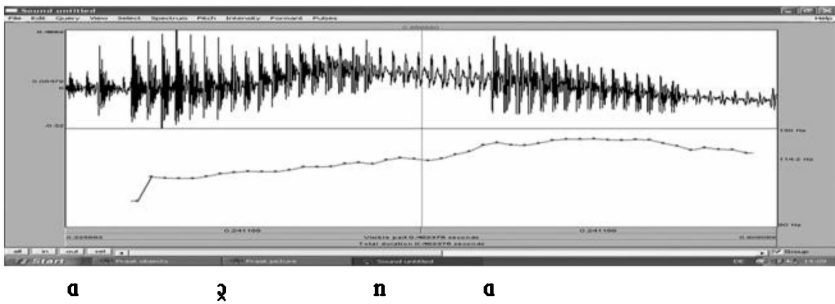


Diagram 2.13: *a = na* [remain=INF]

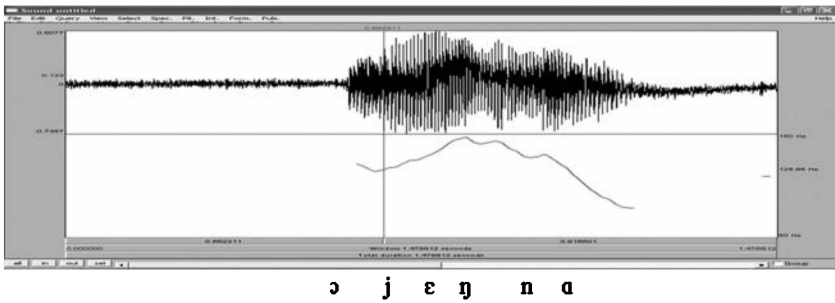


Diagram 2.14: *o - y - en = na* [CAUS-y-return=INF] 'return (TR)'

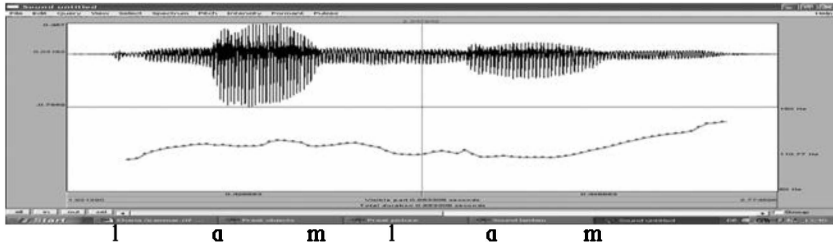


Diagram 2.15: *lam-lam* [seek-RDP] ‘sought for, hunt’¹¹

the infinitival marker =*na* combines with the monosyllabic *aw* ‘remain’ in an LH-pitch pattern, which falls slightly towards the end, but only after having peaked in [nɑ].

As Diagram 2.14 shows, this is not due to any inherent properties of =*na* ‘INF’, as this element does not incorporate into the LH-pattern if the lexical morpheme is bisyllabic. Diagram 2.14 also shows that derivational affixes, in this case the causative marker *o-*, are integrated into the LH pattern.

Just as Diagram 2.14 shows that derivational marking in Kharia is affixal and a part of the prosodic word, we should expect this to be true of other derivational processes as well, including reduplication in the masdar (6.6.2.1). As Diagrams 2.15 and 2.16 show, this is clearly the case. Diagram 2.16 incidentally also shows that the reciprocal marker *kol* is a separate phonological word.

It is not always the case that enclitics have a falling pitch. For example, in Diagram 2.17 we see that the pitch continues to rise gradually throughout both the contentive morpheme as well as the enclitics.

It may even be the case that we find an LH pattern among enclitic elements when more than two combine. This is demonstrated in Diagram 2.18.

At the moment, the data are not sufficient to determine the exact distributional patterns of pitch with multiple enclitics, and it may well be that with two or more enclitic elements, these may combine to form “clitic words” (cf. Aikhenvald, 2002). Although this is rather speculative at present, examples such as Diagram 2.18 at least suggest this possibility.

¹¹ The fact that each *lam* in Diagram 2.15 and each *ol* in Diagram 2.16 retain their own LH-pattern, although they are incorporated into the larger, overall LH-pattern, suggests that these are (phonological) compounds. This topic requires further study.

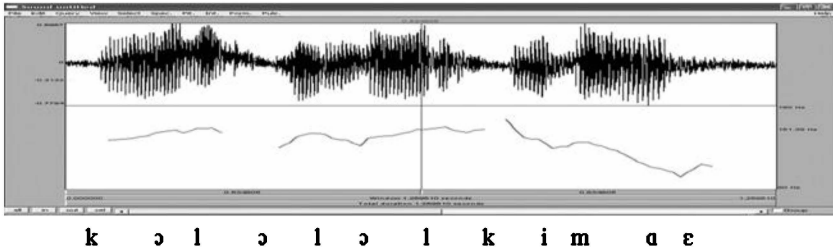


Diagram 2.16: *kol ol-ol=ki=may* [REC take-RDP=MID.PST=3PL] 'they kept on taking each other'

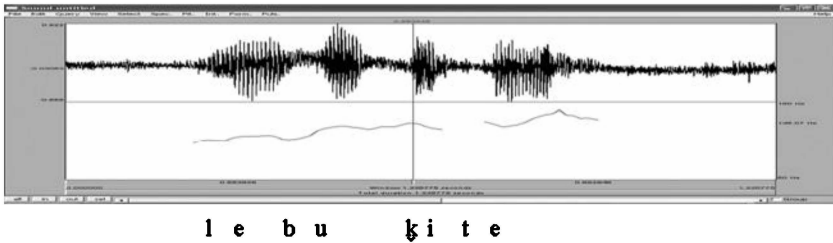


Diagram 2.17: *lebu=ki=te* [man=PL=OBL] 'the men (OBJECT)'

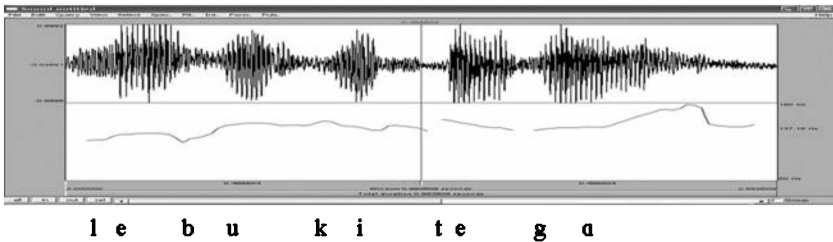


Diagram 2.18: *lebu=ki=te=ga* [man=PL=OBL=FOC] 'the men (FOCUSED OBJECT)'

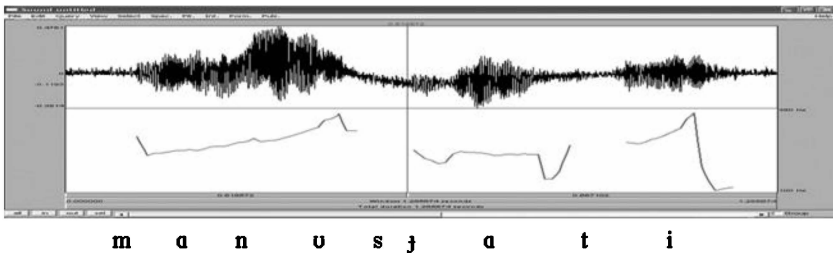
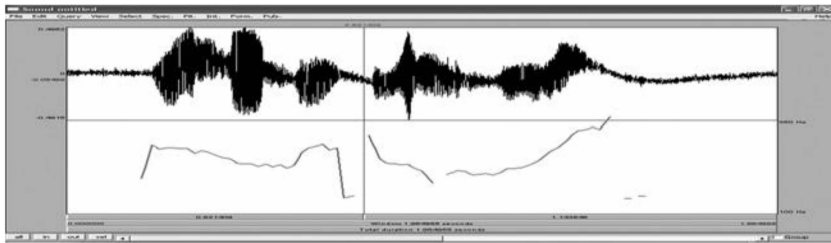


Diagram 2.19: *manus jati* [man caste] 'humanity'



h o l e b u k i y a ? ^ b u ŋ

Diagram 2.20: *ho lebu=ki=ya? buŋ* [that man=PL=GEN with] ‘with those men’

Consider now Diagram 2.19, which in most previous analyses of Kharia would be considered a compound.

Phonologically speaking, there is no reason to consider *manus jati* a compound, as both contentive morphemes are phonological words, unlike *lam-lam* in Diagram 2.15 above. Compounds will be dealt with in further detail in Section 4.6.¹²

Finally, postpositions also seem to qualify as phonological words, since they also show the typical LH-pattern; cf. *buŋ* ‘INST’ in Diagram 2.20.

The status of the demonstratives, on the other hand, is uncertain: Some examples, such as *ho* in Diagram 2.20, allow no conclusions whereas others, at least when preceding a bisyllabic unit with its own LH-pitch pattern, also show an LH-pitch pattern and would seem to constitute a separate word.

Their status is further complicated by the fact that the floating pragmatic enclitics (cf. 3.1.1) can never intervene between a demonstrative and a following unit. This clearly suggests that the demonstratives are proclitic. Furthermore, if a demonstrative precedes a monosyllabic unit such as *kaŋ* ‘person; SG.HUM’, *ghay* ‘way’, *tiʔj* ‘side’ or *jeʔ* ‘SG.NHUM’, it can form a phonological word with this second unit, i.e., *ho=kaŋ* [that=SG.HUM] ‘s/he’, *ho=ghay* [that=way] ‘thus, like that’, *ho=tiʔj* [that=side] ‘to there’ and *ho=jeʔ* [that=SG.NHUM] ‘it’ are all pronounced with an LH-pitch pattern.¹³ Again, this suggests that the demonstratives are proclitic.

¹² Rehberg (2003: 31f.) also comes to a similar conclusion with respect to such units, although her argumentation is one of primary and secondary accents, i.e., in her terms there is apparently neither a primary nor a secondary accent (= low-tone pitch) here, which one might expect in a compound. In my terms, both units show the usual LH-pitch pattern within the same pitch range, with no reason to assume that the two form a tighter phonological unit.

¹³ Alternatively, it could be argued that *kaŋ*, *ghay* and *tiʔj* are enclitic. This issue requires further study.

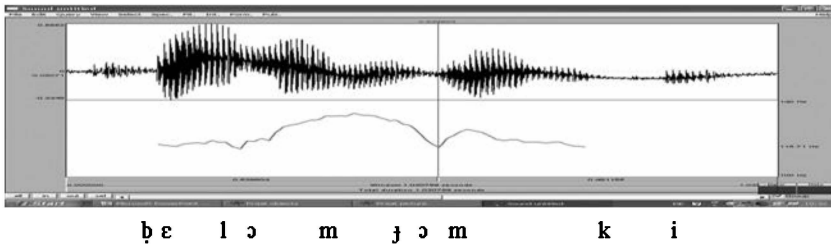


Diagram 2.21: *belom jom=ki* [ripen AUTOPOES=MID.PST] 'it just ripened'

There is, however, evidence suggesting that the demonstratives can, under certain circumstances at least, have the status of independent phonological words, although this is a highly marked construction. Cf. e.g. examples (165) and (166) in Section 5.8.

The status of these elements requires further research. Following Kharia convention, the demonstratives will not be conjoined to the following unit by the sign “=” in this work unless this is a unit which always shows the typical LH-pitch pattern.

There is one further area in which the prosodic status of the respective elements is unclear: the “v2s”. These are grammatical units which have evolved from contentive morphemes, some of which are still homophonous with these morphemes (cf. 6.5). These markers express categories such as the passive voice, reflexivity and *Aktionsart*.

Often, this unit appears to have its own LH-pitch pattern, suggesting that it is a phonological word, as with *jom* in Diagram 2.21.¹⁴

If such a marker follows a monosyllabic contentive morpheme which does not end in /ʔ/, this marker can be incorporated into the LH-pitch pattern, cf. Diagram 2.22.

However, as Diagram 2.23 shows, the v2 can also have a falling pitch, suggesting that it is not a phonological word, although it is also not a part of the same phonological word as *ɣo=ga* in this example.

The pitch on the element *ɔm* does not rise at all in Diagram 2.23, and this element would thus seem to be an enclitic. However, this analysis is problematic, as the v2 *ɔm* here is separated from the semantic base of the predicate by the floating pragmatic clitic =*ga*, which is otherwise always the last unit of the phonological word. More research is necessary on this topic.

¹⁴ Assuming in this case that the pitch on *jom* here falls after rising due to the continually decreasing pitch typically found in predicates after the semantic base.

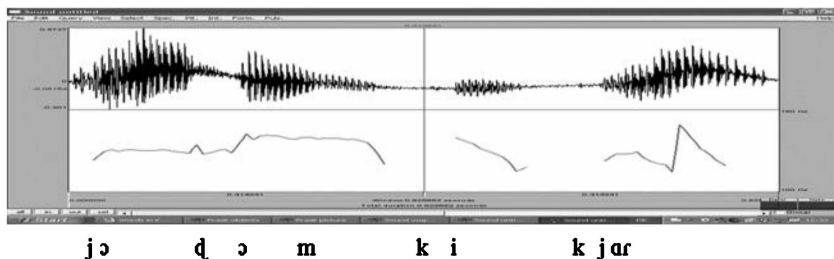


Diagram 2.22: *yo dom=ki=kiyar* [see PASS=MID.PST=3DU] 'they (DU) were seen'

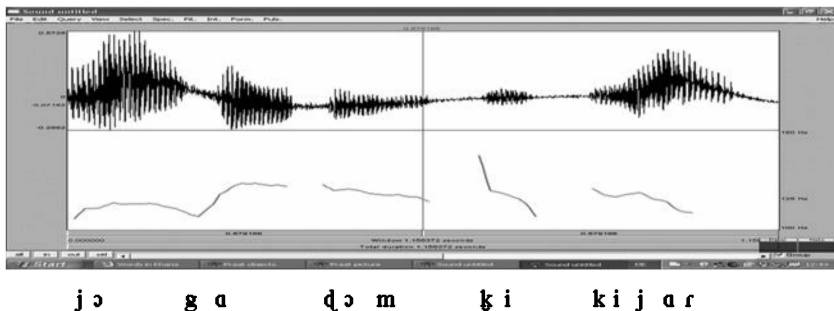


Diagram 2.23: *yo=ga dom=ki=kiyar* [see=FOC PASS=MID.PST=3DU] 'they (DU) were SEEN'

2.6 SENTENCE PROSODY

Little work has been done to date on sentential prosody in Kharia. Rehberg (2003: 37ff.) has a few preliminary remarks on the topic, although she cautiously refers to these as speculative.

With respect to declaratives, Rehberg notes that the utterance is relatively monotonous and decreases gradually both in intensity and frequency towards the end of the sentence. Diagram 2.24 presents a typical example of this from our data, with pitch indicated in the lower half of the diagram.

8. *kahani lebu=ki khorj=ki=te kayom=ta=ki.*
 story person=PL village=PL=OBL speak=MID.PRS=PL
 'The people tell [this] story in the villages.'

With regard to pitch, note that pitch level gradually decreases throughout the utterance, and towards the end of the sentence overrides the LH-



Diagram 2.24

pattern we would expect from *kayom* in the predicate if this were spoken in isolation. Similarly, the intensity (not shown here) also gradually decreases. It is highest on the syllable [bu] in *lebu*, only slightly more intense than on [ha] in *kahani*. After *lebu*, however, it continues to gradually decrease, along with pitch.

A similar intonation pattern is found in the next line from the same story, given here as Diagram 2.25. Here, the speaker is beginning his narrative.

9. *moj raja rani aw=ki=kiyar.*
 one king queen QUAL=MID.PST=DU
 'There were a king and queen.'

In the next line of the story, we observe that the speaker signals at the end of the first clause that his utterance is not yet finished by raising his voice somewhat. For reasons of space, this utterance must be divided into two diagrams. In Diagram 2.26, we see the non-final clause, which rises towards the end, followed by a pause (0.565sec).

This is then followed by the second half of the utterance, shown in Diagram 2.27, with falling intonation towards the end and relatively monotonous otherwise.

10. *raja=ya? jimi aw=ki sembho, oqo? rani=ya? jimi*
 king=GEN name QUAL=MID.PST Sembho and queen=GEN name
aw=ki dakay.
 QUAL=MID.PST Dakay
 'The king's name was Sembho, and the queen's name was Dakay.'

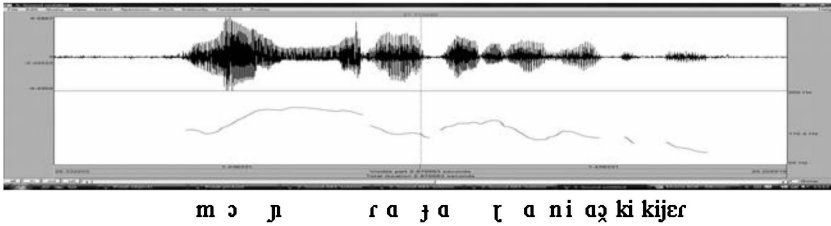


Diagram 2.25

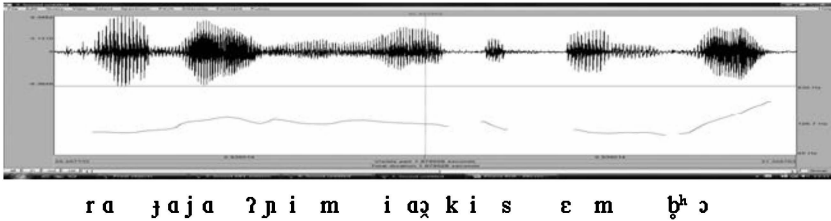


Diagram 2.26

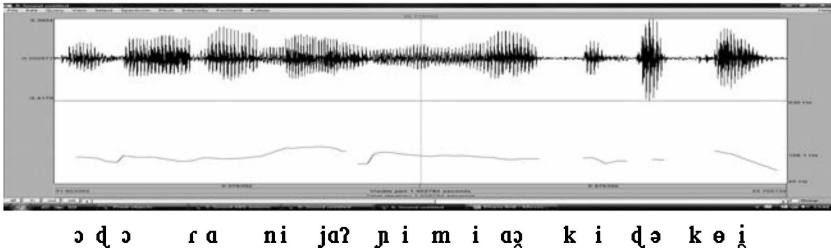


Diagram 2.27

In interrogatives, on the other hand, Rehberg notes that the pitch rises and falls considerably in the utterance, and each low-tone pitch (marking the beginning of a phonological word) is clearly articulated. Sentence-pitch only falls towards the end of the utterance. Unfortunately, due to the nature of the corpus, there are only very few interrogatives found in our corpus, as most texts were either written texts or narratives, and most of these are constituent questions, i.e., there are few polar questions, and the few that are found are to a certain extent “artificial”, as the speaker is playing the part of the person posing the question. Nevertheless, Rehberg’s observation seems to be confirmed to some degree by these data, although it is difficult to consider the low-tone pitch here to be especially

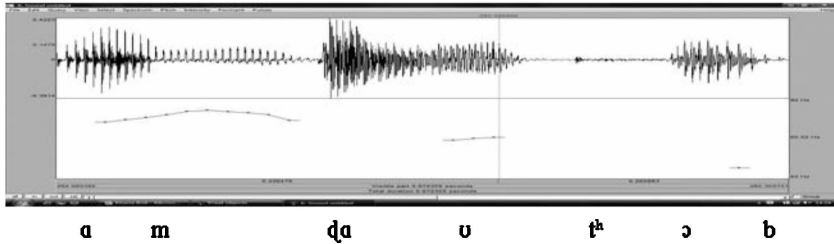


Diagram 2.28

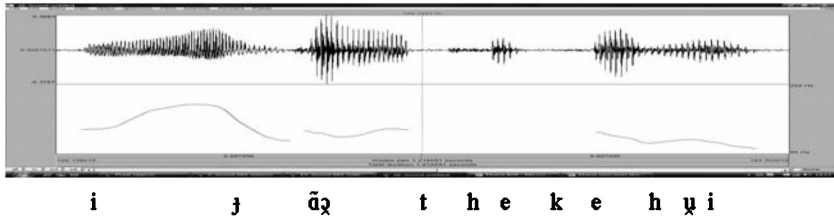


Diagram 2.29

clearly marked, although this may also be due to the quality of our recording. Consider first Diagram 2.28, one of the rare polar questions in our corpus.

11. *am dʌ[ʔ] uʰ=ɔb?*
 2SG water drink=ACT.PST
 'Did you drink water?'

A similar pattern is found in constituent questions, as shown in Diagram 2.29, although the final drop in intonation seems to be less sudden.

12. *i jãʁ heke hoy?*
 what animal QUAL INFER
 'What animal could it be?'

Much research is still needed in this area, especially with more data from spontaneous spoken language, before any further conclusions can be drawn on sentential prosody.

SYNTACTIC ATOMS

As Dixon & Aikhenvald (2002a: 2) note, not all languages have a concept with (more or less) the same meaning as the English concept *word*. In a certain sense this is also true of Kharia: While there is the concept *sabda* with (roughly) the same meaning as English *word*, this term has been borrowed from Indo-Aryan, cf. Sadri *sabad* and Hindi *śabda*, both of which mean approximately the same thing as English *word*. The closest native term in Kharia is *kayom*, which is perhaps best translated as ‘speech; matter; (to) speak; spoken’. Thus what this element primarily refers to is the act of speaking in general and it can refer to entire utterances or, for that matter, “matters” in general. It does not, however, refer to the same unit as *word*, regardless of how this is defined.

It was shown in Chapter 2 that a phonological word in Kharia typically consists of a (generally bisyllabic) contentive morpheme and potentially one or more enclitics. If there are two or more enclitics appearing together, there is preliminary evidence that these may combine to form a phonological word, although this is still somewhat tentative.

The following section presents syntactic evidence which demonstrates that those elements which were shown to behave as clitics phonologically in Section 2.5, in that they do not show the LH pitch pattern, also behave syntactically as enclitics in that they attach to entire syntactic units, not lexical roots or stems. As such, these units—and their hosts—both qualify as “syntactic atoms” (Di Sciullo & Williams, 1987) or “grammatical words” (Dixon & Aikhenvald, 2002a) in Kharia. Thus, phonological words in Kharia often consist of more than one “grammatical word” or “syntactic atom”

3.1 CLITICS AS “PHRASAL AFFIXES”

There seems to be almost unanimous agreement in studies on clitics that, from a morphosyntactic point of view, these are “words” or syntactic-level units. For example, Sadock (1991: 52) includes this as one of the typical characteristics of clitics: “Clitics are independent elements of syntax.” A similar view is found in Aikhenvald (2002: 54): “The mismatches

between grammatical and phonological words. are most often accountable for by clitics which form a phonological word with their host, but can be considered a grammatical word in their own right.” Similarly, Dixon & Aikhenvald (2002a: 25) write “The term clitic is often used to refer to something that is a grammatical word but not a complete phonological word (for example, it does not take stress).” Anderson (2005), on the other hand, differentiates clearly between phonology and syntax; he speaks of “phonological clitics”, units “whose phonological form is deficient in that it lacks prosodic structure at the level of the (Prosodic) Word” (Anderson, 2005: 23), but also “morphosyntactic clitics”, “whose position with respect to the other elements of the phrase or clause follows a distinct set of principles, separate from those of the independently motivated syntax of free elements in the language.” (Anderson, 2005: 31). It is this view of “clitics” which will be taken here, i.e., an element may be clitic in a phonological sense, in a morphosyntactic sense, or in both senses.

As we saw in Section 2.5, there are a number of morphemes in Kharia which may be termed phonological clitics, as they do not have the typical LH pitch pattern found in phonological words. In this section, we are concerned with whether these elements may also be considered morphosyntactic clitics, often referred to as “phrasal affixes”, i.e., grammatical words or syntactic atoms which attach to a possibly complex unit in the syntax (i.e., a “phrase”). We will assume here that any element which is considered a (morphosyntactic) clitic, as opposed to an affix, will necessarily attach to a host which is itself a syntactic atom. Although this does not rule out cases in which a host consists of a simple contentive morpheme, the crux of the matter is whether or not the marker in question CAN take a “phrasal” host or whether it is restricted to hosts which are lexical “roots” or “stems”

We begin with a discussion of the markers of functional categories such as pragmatic marking, inalienable possession, number, TAM, case and person, where the clause-level relevance of the information marked by these forms is obvious. The case is somewhat different with the genitive, which is not relevant at the clause level in the sense that it does not serve to mark the function of a TAM/PERSON-syntagma (“verb”) or a Case-syntagma (“NP”) in the clause as, e.g., subject, object, etc. Nonetheless, as we shall see, the genitive attaches to syntactically relevant units, such as “phrasal” semantic bases, which may consist of a demonstrative, quantifier and possibly one or more contentive morphemes (cf. Chapter 4), integrating what could serve as a (possibly complex) semantic base of a TAM/PERSON-syntagma or Case-syntagma into a larger unit. As such, it must be considered a syntactically relevant unit.

3.1.1 *The pragmatic markers*

There are three primary pragmatic markers in Kharia, the general focus marker =*ga*, the contrastive focus marker =*ko* and the additive focus marker =*jo*.¹ These markers may be termed “floating clitics” as they may attach to any element of the clause provided that this unit is a phonological word, with the sole exception of the demonstratives (whose status as proclitics / phonological words is uncertain), although they are usually found at the end of a phrase. The only restriction is that they appear at the end of some unit that can stand alone as a phonological word. The following provides a few examples:

1. *ho rusuŋ oʔ=te=ga* ‘in that red house’; ‘that red house (OBJECT)’
 that red house=OBL=FOC

Here, the scope of the focus is indeterminant: The focus domain may either be the entire Case-syntagma (e.g., *that red house* and not *the blue hut*) or simply the last element (e.g., *house* as opposed to *hut*). In the following example, this indeterminancy is not found. SMALL CAPS indicate the intended focus:

2. *ho rusuŋ=ga oʔ=te* ‘in that RED house’; ‘that RED house (OBJECT)’
 that red=FOC house

However, as noted above, the demonstrative may not be separated from a following lexical unit in the same phrase:

3. **ho=ga rusuŋ oʔ=te* ‘in THAT red house’; ‘THAT red house (OBJECT)’
 that=FOC red house

The following examples further demonstrate the possible distribution of these pragmatic markers in various complex units:

4. *karay goʔd̥=te=ga* / *karay=ga goʔd̥=te* ‘s/he does the job’
 do C: TEL=ACT.PRS=FOC do=FOC C: TEL=ACT.PRS
5. *ek t̥ho=ga* ‘one’ / *ek=ga t̥ho(=ga)* ‘one’
 one CLASS=FOC one=FOC CLASS=FOC

¹ For a discussion of their respective pragmatic functions, see Section 7.7.4.

6. $u=je[ʔ]=ko$ $ho=ki=yaʔ=ga$ *heke*.
 this=SG.NHUM=CNTR that=PL=GEN=FOC QUAL.PRS
um=ijn *ter=e*.
 NEG=1SG give=ACT.IRR
 ‘BUT THIS IS THEIRS. I will not give [it to you].’ [TK, 2:43]
7. *kulam=ki aloy=na=pe* *musa=ko yesu jorme=ki*.
 brother=PL sing=MID.IRR=2PL today=CNTR Jesus be.born=MID.PST
 ‘Sing, brothers, FOR TODAY Jesus is born.’ [RK, 8:2]
8. *ho=kaʔ* *qisaʔ qisaʔ* *khoʔtay=jo* *merom* *gupa=na*
 that=SG.HUM far REP up.to=ADD goat guard=INF
laʔ=khoʔ.
 IPFV=PST.II
 ‘He used to tend the goats, EVEN GOING (= ALSO UP TO) VERY FAR.’ [RD, 1:4]
9. *tay* *adj=te=jo* *batay=oʔ* *no* “*dhǎy[=e]* *am, han*
 then ANAPH=OBL=ADD tell=ACT.PST Cmpl hurry=ACT.IRR 2SG that
maʔa boʔ=te *qa[ʔ]* *[a]yiʔj*. *uq* *qe=na=m*.”
 cave place=OBL water QUAL.PRS drink come=MID.IRR=2SG
 ‘Then he said TO HIM, TOO, “Hurry, at the cave there is water. Drink and come back” ’ [AK, 1:38]

Two or all three of these floating enclitics may also appear together, apparently in any order, although there are slight differences in meaning, depending on the particular scopal relations. Unfortunately, it was not possible to get at these semantic differences during interviews, although most speakers felt there were slight differences in emphasis. Thus, the following all have approximately the same meaning:

10. $am=te=jo=ko$ / $am=te=ko=ga$ / $am=te=ga=ko$ / $am=te=ga=jo$
 2SG=OBL=ADD=CNTR =FOC
yo=yoʔj.
 see=ACT.PST.1SG
 ‘I saw YOU (as well).’

These three units may thus safely be classified as enclitics according to their distribution, which fits in well with their prosodic properties, i.e., no inherent LH pitch pattern.

3.1.2 *Case and number marking*

The most obvious case-marking candidate for enclitic status as opposed to suffixal status in Kharia is the genitive marker //YAʔ//. As the following shows, similar to 's in English the genitive marker in Kharia attaches to its host at the syntactic level.

11. *laʔ u sembho ro ɖakay rani=kiyar=aʔ nãw jan*
 then this Sembho and Dakay queen=DU=GEN nine CLASS

beʔt=qom=kiyar² aw=ki=kiyar.

SON=3POSS=HON QUAL=MID.PST=HON

'And this Sembho and queen Dakay had nine sons (HON) (= *This Sembho and Queen Dakay's* nine sons were).'

12. *ayo apa=yaʔ jiyom=te=ko kundʉʔ=ki=yaʔ thoŋ soub*
 mother father=GEN soul=OBL=CNTR child=PL=GEN for all

bhere=ga dular aw=ta.

time=FOC love QUAL=MID.PRS

'In a mother or father's soul there is always love for their children.'

[RD, 2:110]

The unit to which the genitive marker attaches may of course consist of a single contentive morpheme, e.g. *lebu=yaʔ* [person=GEN] 'of the person; the person's', but this is merely a special case in which the syntactic unit simply happens to consist of only a single morpheme.

Similarly, the oblique marker =*te* and the plural marker =*ki* are enclitic forms which attach to the last element of a syntactic construction, regardless of the status of this element. Consider first =*ki* 'PL' in following example:

13. *munuʔsiŋ rochob=aʔ lebu=ki=ko*
 east side=GEN person=PL=CNTR
 'the people of the east; the easterners'

² Although the form *beʔt=qom=kiyar* would be expected due to the phonological system of the language (cf. Section 2.2.2), the morpheme for 'son' is consistently written as *beʔt* or *beʔ* as it derives from Sadri *beʔa*.

At first sight it might seem reasonable to assume that =*ki* is a suffix denoting number, which combines with the lexical root/stem *lebu* to form a syntactic atom.

There is, however, convincing evidence that this is not the case. Note that in Kharia it is possible to omit virtually any lexical information which is considered obvious from the context. Generally this affects entire syntactic units, as for example when number and case-marked arguments are not explicitly stated as their identity is obvious (see 7.1.1). However, this same principle may also apply only to the semantic head of an argument, while the final case/number marking is retained. Here the lack of an explicitly mentioned semantic head fills a function similar to the use of *one* in the English *Which hat did you buy? The green one*. In Kharia, no proform need be used and the contentive morpheme is simply not explicitly mentioned.

In cases such as these, the grammatical markers which would otherwise directly follow the semantic head (e.g., possessive, number or case marking) now attach to the last element of the syntactic construction without the lexical head, regardless of the status of this element. Thus, (13) above can also be realized as in the following, attested form, whose grammaticality was also confirmed by other speakers.

14. *mumu?siŋ rochob=a?=ki=ko*
 east side=GEN=PL=CNTR
 ‘the easterners; the ones of the east’ [HJPa:120, line 29]

Similar data are found for the oblique marker =*te*. In the following two examples, =*te* attaches to the otherwise final element of the semantic base, regardless of its status, when the semantic head, *lebu* in the first example and *bo?* in the second, is not explicitly mentioned:

15. *ho lebu=ki=te* ‘those people (OBJ) or simply
 that person=PL=OBL
ho=ki=te ‘them, those ones’ (OBJ)
 that=PL=OBL
16. *ijŋ=a? bo?=te* or simply *ijŋ=a?=te* ‘at my place’
 1SG=GEN place=OBL 1SG=GEN=OBL

One final example from our corpus is presented below, in which =*te* also has scope over the entire (non-italic) phrase.

17. *pheinga*, *am=ko* *ku²j=na* *ro* *ebo²=na=te=ga* *samay*
 grasshopper 2SG=CNTR dance=INF and play=INF=OBL=FOC time
bitay=ob. *a²b* *hawa* *jog=e!*
 spend.time=ACT.PST.2SG now air eat=ACT.IRR
 ‘Grasshopper, you spent your time *in dancing and singing*. Now eat
 air!’ [TK, 2:45]

On the basis of data such as these, we conclude that case and number markers are “phrasal affixes” This is compatible with the phonological data presented in Chapter 2, where it was shown that these markers are phonologically not integrated into the LH pitch pattern of the word, except under very special conditions. They may thus be considered enclitics from a syntactic perspective as well.

3.1.3 Markers for inalienable possession

Similar to the data above, markers for inalienable possession can be shown to attach to syntactic units, even if these generally consist of a single word. In the following example, the possessive marker =*dom* has scope only over the word to which it is directly attached and both entities are marked individually:

18. *kole²=ya²* *ma=dom* *ro* *apa=dom* *ho=kiyar=te* *jal* *tay*
 parrot=GEN mother=3POSS and father=3POSS that=DU=OBL net ABL
oj=na *thon* *han=te* *u=te* *pu²d-pu²d*
 take.out=INF PURP that(DIST)=OBL this=OBL jump-RDP
khora=ta²j=kiyar *ro* *sob=ga* “*tāy tāy*” *toro²d=taj=ki*.
 ITER=MID.PROG=3DU and all=FOC “*tay, tay!*” cry=MID.PROG=PL
 ‘The [two] parrots’ *mother* and *father* kept jumping around here and there to free them from the net and all [the birds] are crying “*Tay! Tay!*” ’ [BB, 2:34]

Nevertheless, as the following two examples show, markers for inalienable possession in fact attach to entire “phrases” and have scope over these:

19. *bhala* ayo aba=dom=ko *sadhu je? kunḍu? ho=ghay.*
 well mother father=3POSS=CNTR simple FOC child that=way
 ‘Well, *his mother and father* were simple, like children.’ [RD, 2:6]
20. *modi ro saw-ṭay=dom* ayo aba ro boker
 Modi and spouse-woman=3Poss mother father and brother.in.law
kulam=dom=ki=ya? kaṭa sumbo?=te gur=kon ṭyam ṭyam
 brother=3POSS=PL=GEN foot base=OBL fall=SEQ cry REP
maph bor=o?=kiyar.
 forgiveness ask=ACT.PST=DU
 ‘Modi and his wife fell down at the feet of *his mother, father, brothers-in-law and brothers* and, crying, asked for forgiveness.’ [RD, 2:113]

3.1.4 Markers for TAM and person / number / honorific status

Turning our attention now to the functional markers found in TAM/PERSON-syntagmas (“verbs”), the most clearly enclitic markers are those of person/number/honorific status: Generally speaking, this marking in non-negated TAM/PERSON-syntagmas attaches to the last element of this unit whereas with negated TAM/PERSON-syntagmas, the negative marker *um* appears directly before the semantic base and person marking attaches to this marker (cf. 6.7.1):

21. *kayom=ta=ṅ* *um=iṅ* *kayom=ta*
 speak=MID.PRS=1SG NEG=1SG speak=MID.PRS
 ‘I speak’ ‘I do not speak’

However, in the second person singular, person marking with negated TAM/PERSON-syntagmas can attach either to the negative marker *um* or to the last element of this unit. Cf. (from Malhotra, 1982: 285):

22. *ubhroṅ* *um=em* *ḍe=na* or
 these.days NEG=2SG come=MID.IRR

ubhroṅ *um* *ḍe=na=m.*
 these.days NEG come=MID.IRR=2SG
 ‘These days you do not come.’

Furthermore, with cosubordination (6.6.1), the subject marking has scope over more than one predicate base marked for TAM:

23. *ɲokh=o?* *uʰh=o?=ki*.
 eat=ACT.PST drink=ACT.PST=PL
 ‘They ate and drank.’

The analysis of person marking as a morphosyntactic enclitic is thus the most natural analysis: Although this marking is subject to (relatively) strict rules regarding its position with respect to the TAM/PERSON-syntagma, the fact that it does not form a very tight unit with the remaining TAM/PERSON-syntagma shows that it is not a suffix. Note that this analysis will also force us to analyze this marker as the subject itself and not merely an “agreement marker”, as it is a syntactic unit. This will be of importance in the discussion in Section 7.1.2.

The TAM/basic voice markers can also be shown to be “phrasal affixes” For example, the marker of the past tense, middle voice, =*ki* and its active counterpart, =*o?*, would at first sight appear to be suffixes which attach in the morphology to a lexical morpheme, e.g. *col-ki=may* [go-MID.PST=3PL] ‘they went’ or *karay-o?=may* [do-ACT.PST=3PL] ‘they did’ However, consider the following examples:

24. *ho rochoʰb=ki=ɲ*.
 that side=MID.PST=1SG
 ‘I moved to that side.’

25. a. *ho=kaɾ* *ɲɲ=a?* *natgot=ki*.
 that=SG.HUM 1SG=GEN family.relationship=MID.PST
 ‘S/he became a member of my family.’

- b. *ho=kaɾ=te* *ɲɲ=a?* *natgotʰ=oʰj*.
 that=SG.HUM=OBL 1SG=GEN familyrelationship=ACT.PST.1SG
 ‘I accepted him/her as a member of my family.’

The decisive issue in these examples is that if these grammatical markers attach to lexical stems as suffixes, the structure of these TAM/PERSON-syntagmas becomes exceedingly difficult to interpret. For example if we analyze =*ki* in (24) as a tense suffix (ignoring here the person marker =*ɲ*, which as we have seen is enclitic), this would presumably make *rochoʰb* a verb, as this is the “root” or “stem” it attaches to. The question then arises as to the status of *ho* ‘that’, which is clearly a modifier of *rochoʰb* ‘side’ Similar comments hold in (25) for *ɲɲ=a?* and its relation to *natgot* if =*ki* and =*o?* are analyzed as suffixes.

It will also not be feasible to analyze these structures as “compounds” or what Di Sciullo & Williams (1987) refer to as “syntactic objects” These latter units are X^0 s or “leaves” in the constituent structure, but of a different nature than most syntactic atoms, as they are formed in the syntax and then reinterpreted as syntactic atoms, such as *essui-glace* ‘windshield wiper’ in French (see the discussion in Di Sciullo & Williams, 1987, chapter 4). However, unlike the syntactic objects of French, these units in Kharia are not opaque—e.g. they can contain deictic and anaphoric reference as in (24) and (25)—and they are fully productive.

Rather, their structure can most easily be accounted for if one considers the grammatical markers, here TAM/basic voice markers, as attaching to the entire phrase and not to roots/stems. In other words, only an analysis of these markers as morphosyntactic enclitics is able to account for their distribution. As with the enclitics discussed above, this is once again in line with their prosodic properties, as these units do not show an LH-pitch pattern.

3.1.5 *The perfect marker =siʔ(d) and the infinitive marker =na*

Most examples of the markers for the perfect and infinitive in the data are of the type in the following example, in which they follow—and have scope over—a single contentive morpheme functioning as the semantic base of the predicate:

26. *souʔb* *apan* *apan* *muʔ=sɪʔ=may*, *ɖaʔ* *lam=na*.
 all REFL REP emerge=PERF=3PL water search=INF
 ‘All *have set out* on their own, *to look* for water.’

There are, however, a number of examples which confirm the status of these markers as “phrasal affixes”:

Perfect

27. *janʔkoy* *puja* *ɖinu* *ɖam* *ɖel=sikh=oʔ*.
 spring.festival worship day arrive come=PERF=ACT.PST
 ‘The day [for doing] *Jankoy puja* [= a type of religious ceremony] *had arrived*.’
 [MT, 1:126]

28. *muda musa* *boiri=ki* *ɖel* *ɖam=sɪʔ=may*...
 but today enemy=PL come arrive=PERF=3PL
 ‘But today the enemies *have arrived*’
 [Kerkettā, 1990: 4]

29. ... *sosreir=ga?* rema? d̪aŋ=si?=may, ...
 husband's.parents'.home=FOC call send=PERF=3PL
 the grandparents *have called [him] and sent [him] back,*

Infinitive

30. *u=kɑr=a?* thoŋ buɽha=ki=te ikuʔd̪ ikuʔd̪ dhain
 this=SG.HUM=GEN for old.man=PL=OBL very REP thank
 mane=na ayiʔj.
 honor=INF QUAL.PRS
 ‘For this [= for what they did], the elders should be thanked and honored (= it is *to thank [and] honour* the elders for this).’ [MT, 1:183]

31. *ho=ki=te u raij tay ɲon d̪aŋ=na*
 3=PL=OBL this kingdom ABL drive send=INF
 ayiʔj.
 QUAL.PRS
 ‘[We] have *to drive them out* of this country.’ [Kerkeṭṭā, 1990: 11]

We conclude that the both the perfect and the infinitival markers are “phrasal affixes” and that they are therefore syntactic atoms, despite the fact that they are not phonological words. Both attach to their host at the “phrasal” level—although this “phrase” often consists of a single contentive morpheme. This fits in well with their phonological characteristics, in that these two markers do not integrate into the LH pitch pattern of the prosodic word, except under very special conditions.

3.2 CLITICS IN KHARIA—AN OVERVIEW OF THEIR CHARACTERISTICS

Having shown in Sections 2.5 and 3.1 that the grammatical markers discussed in those sections are both phonologically and morphosyntactically enclitic, in this section these clitics are classified primarily according to the criteria given in Aikhenvald (2002).

3.2.1 *Position with respect to host*

With the exception of units such as the demonstratives, whose clitic status is somewhat uncertain, all other clitics in Kharia attach to the end of their respective hosts. Where more than one clitic attaches to a particular unit,

the enclitics appear in a fixed order, with the exception of the floating clitics, which may appear in any order with respect to one another.

Person marking on the TAM/PERSON-syntagma is a further slight exception to this rule. As shown Section 3.1.4, in non-negated TAM/PERSON-syntagmas, person marking is always the final marker on the TAM/PERSON-syntagma (unless one of the floating clitics also marks this unit). However, if the TAM/PERSON-syntagma is negated, these markers appear before the remaining unit, following the negative marker *um*, which precedes the semantic base. These are then what are commonly referred to as anticipatory clitics.³ Thus, person markers in Kharia are either enclitic or anticipatory clitic.

3.2.2 *Segmental properties of clitics and their respective degree of cohesion*

3.2.2.1 *Segmental properties*

With respect to the segmental and phonotactic properties of clitics compared with other morpheme types, we find tendentially different structures for the two categories but no hard-and-fast rules. Contentive morphemes tend to be either bisyllabic with the structure (C)VC(C)V(C) or, if monosyllabic, to have the segmental structure (C)VC. The following provides a few examples:

32. Bisyllabic: *kayom* ‘speech; matter; speak’, *endu?* ‘look at, stare’
 Monosyllabic: *mo²d* ‘eye’, *ro?* ‘spill’, *o?* ‘house’

Similarly, grammatical morphemes which derive from contentive morphemes, such as “v2s” (6.5) tend to have the same structure—in general they are monosyllabic and have the form CVC:

33. *dom* ‘PASS/REFL’, *go²d* ‘C:TEL’, *do²d* ‘A.TEL’

³ Anticipatory clitics are much more common in the North Munda languages such as Santali and Mundari, as person marking there generally appears on the last element preceding the predicate—whether negated or not—unless the predicate is the only element in the clause, in which case it is enclitic to the predicate.

Clitics, on the other hand, are almost always monosyllabic and have the form (C)V:

34. =*ki* 'MID.PST', =*te* 'OBL', =*e* 'ACT.IRR'

Nevertheless, this is only tendentially so. For example, there are a number of clitics with the structure (C)VC and one bisyllabic morphosyntactic clitic.

35. =*siʔ* / *siʔd* 'PERF', =*khoʔ* 'PST.II', =*nom* '2SG.POSS', =*niŋ* '1PL.INCL',
= *kiyar* 'DU'

We also find a few monosyllabic contentive morphemes and "v2s" with the structure (C)V, otherwise typical of clitics:

36. *yo* 'see', *i* 'what?', *ʔu* 'DPT'

Thus, the segmental structure of enclitics differs only tendentially from that of other morphemes and their segmental structure is not a definitional criterion for their enclitic status.

3.2.2.2 Degree of cohesion with host

Although clitics generally attach to their hosts without further phonotactic changes, a number of clitics do either induce a change in the form of their host or are themselves subject to phonotactic changes due to either preceding or following segments:

- =*oʔ* 'ACT.PST' is the only clitic which causes a change in the form of its host. A consonant preceding =*oʔ* must be voiceless and aspirated. /ʔ/, from underlying /g/ (2.2), is realized as /kh/: *ʔoʔ* 'eat', *ʔokh=oʔ* 's/he ate', *sanɡoʔd* 'close' *sanɡoʔh=oʔ* 's/he closed' After vowels, the form /yoʔ/ is found: *yo* 'see', *yo=yoʔ* 's/he saw' Some younger speakers, however, tend to use /yoʔ/ in all environments, including after consonants, e.g., *remaʔ* 'call', *remaʔ=yoʔ* 's/he called', although its use after consonants is rather seldom and the form *remakh=oʔ* 's/he called' is more common.

=*oʔ* also affects some following units: In combination with the person markers of the first and second persons, singular, /iŋ/ and /em/, respectively, the resultant forms are =(y)oʔj and =(y)oʔb.

- =*siʔ* / *siʔd* can optionally be realized as =*chiʔ* / =*chiʔd* following a morpheme-final /j/: *gaʔj=chiʔd=iɲ* or *gaʔj=siʔd=iɲ* ‘I have cut’ from *gaʔj* ‘cut’
- The form of the genitive marker //YAʔ// depends largely on the final segment of the preceding syllable. Although younger speakers tend to use the form =*yaʔ* rather consistently regardless of the final segment of the preceding syllable, the general form of this marker is /aʔ/ following a consonant, /yaʔ/ after /e/ and /i/, /ʔ/ following a syllable ending in /a/, /waʔ/ after /o/ and /u/, /naʔ/ occasionally in demonstratives, and /gaʔ/ in a number of deictic expressions. Whereas these last two forms are lexically conditioned, the remaining forms are strictly phonologically determined. The following presents a few examples:

37. *aniɲ* ‘1PL.INCL’, genitive: *aniɲ=aʔ*
khori ‘village section’, genitive: *khori=yaʔ*
khariya ‘Kharia’, genitive: *khariya=ʔ* / *khariya=yaʔ*
lebu ‘man; person’, genitive: *lebu=waʔ* / *lebu=yaʔ*
u ‘this’, genitive: *u=naʔ* / *u=waʔ*
musa ‘today’, genitive: *mus=gaʔ*

- Clitics beginning with or ending in /e/ or /i/ (often written <y> in syllable-final position) often lose this vowel when the preceding or following elements end in or begin with a vowel. The following presents a few examples:

38. *am=pe* ‘2=2PL’, genitive: *am=p=aʔ*
koɲ=te=jar [know=ACT.PRS=1DU.EXCL] ‘we know’ vs. *koɲ=t=iɲ* [know=ACT.PRS=1SG] ‘I know’
ter=e [give=ACT.IRR] ‘s/he will give’ vs. *batay* (< /batay=e/) ‘s/he will narrate’

- With the middle irrealis marker =*na*, the vowel in the marker of the first person, singular, =*iɲ*, may also be deleted, although this is optional, even for one and the same speaker/writer, as in the following two examples:

39. *am=bar* *leɲ=na=bar* *laʔ* *iɲ=jo* *latke=ga*
 2=2DU fly=MID.IRR=2DU then 1SG=ADD swing=FOC

co=na=iɲ.

go=MID.IRR=1SG

‘You (DU) fly and I will also go swinging.’

[TK, 1:19]

40. *kulu cilay=ki, "ijn um=ijn kayom=na? jarur*
 turtle shout=MID.PST 1SG NEG=1SG speak=MID.IRR certainly
kayom=na=ɲ!
 speak=MID.IRR=1SG
 ‘The turtle shouted, “I shouldn’t speak? Of course *I will speak!*”’
 [TK, 1:36]

- Following vowels, /e/ is generally realized as =*ye*: *i=ye=niŋ?*
 [what=ACT.IRR=1PL.INCL] ‘What shall we do?’

3.2.3 Clitics as “reduced” phonological words

With the exception of the personal markers of the first and second persons, no other clitics can be considered reduced forms of free-standing morphemes. Table 3.1 presents the correspondence between the clitic personal markers and their respective free-standing forms:

Table 3.1: Free-standing and clitic personal markers

	Free-standing	Clitic
Singular		
1	<i>ijn</i>	= <i>(i)ɲ</i>
2	<i>am</i>	= <i>(e)m</i>
Dual		
1 Inclusive	<i>anaŋ</i>	= <i>naŋ</i>
1 Exclusive	<i>ijjar</i>	= <i>jar</i>
2	<i>ambar</i>	= <i>bar</i>
Plural		
1 Inclusive	<i>aniŋ</i>	= <i>niŋ</i>
1 Exclusive	<i>ele</i>	= <i>le</i>
2	<i>ampe</i>	= <i>pe</i>

3.2.4 The relative ordering of clitics

Class I—The innermost layer

Clitics belonging to this class are found in the semantic base of both TAM/PERSON-syntagmas and Case-syntagmas. The following shows their order with respect to one another (“X” refers to the final element of the semantic base, regardless of its status):

41. (DEM=)⁴X=POSS=NUM=GEN

Class II—The medial layer

Clitics belonging to this class are found either only in Case-syntagmas (“NPs”) or only in TAM/PERSON-syntagmas. Members of one group may not be combined with members of the other group:

Class IIa: Clitics of a Case-syntagma (“NP”)

“X” here refers to the entire structure given in (41), which may be marked for any or all of the clitics shown there. Note that this “class” consists only of the oblique marker =te.

42. X=OBL

Class IIb: Clitics of the TAM/PERSON-syntagma (“verb”)

(Here “X” is as in Class IIa but may also be followed by the telic v2s (Section 6.5):

X=PERF=TAM/BASIC VOICE=PERSON MARKING OR X=PERF=INF⁵

Class III: Floating clitics

Finally, as noted above, the floating pragmatic clitics =ga ‘FOC’, =ko ‘CNTR’ and =jo ‘ADD’ may attach to any phonological word or to an entire complex syntactic unit. They may not, however, appear between two clitics or between a clitic and its host.

43. *hoqom=ga=ki=te, *hoqom=ki=ga=te,
other=FOC=PL=OBL other=PL=FOC=OBL

hoqom=ki=te=ga ‘the others’ (focused object)
other=PL=OBL=FOC

44. *ge²b go²d=ga=ki, ge²b=ga go²d=ki,
burn C:TEL=FOC=MID.PST burn=FOC C:TEL=MID.PST

ge²b go²d=ki=ga ‘It burned.’
burn C:TEL=MID.PST=FOC

⁴ Recall from Section 2.5 that the status of the demonstratives as clitics is still rather tentative.

⁵ The infinitive marker is actually an “intermediate” form between Classes II a and b in that it may also be marked for case.

3.2.5 *The varying mobility of clitics*

The above classification of clitics according to position is closely related to their classification according to mobility. To keep the two separate, the groupings will be referred to with respect to mobility as “groups” so as not to confuse these with the “classes” according to position.

Group I: No mobility

In this group we find all of the clitics given above under Classes I and II (a and b), with the exception of person marking. Common to all elements of this group is the fact that their position is fixed, both with respect to other clitics as well as with respect to their host. Summarizing the information above, this group includes the following clitics:

45. (Demonstratives: *u* ‘this’, *ho* ‘that (MED)’, *han* / *hin* ‘that (DIST)’)

Genitive: *a?*

Case: *te* ‘OBL’

Inalienable possession: *na* ‘1(SG)’, *nom* ‘2(SG)’, *dom* ‘3(SG)’

Number: *kiyar* ‘DU’, *ki* ‘PL’

Perfect: *si?* / *si?d*, infinitive: *na*

TAM/basic voice: *ki* ‘MID.PST’, *o?* ‘ACT.PST’

Group II: Low mobility

The only members of this group are the personal markers in the TAM/PERSON-syntagma. Recall from the discussion above that these markers are either enclitic to the non-negated TAM/PERSON-syntagma or are anticipatory clitics (obligatorily, except in the second person, singular) with a negated TAM/PERSON-syntagma:

46. *koŋ=t[e]=iŋ* ‘I know’ *um=iŋ* *koŋ=te* ‘I don’t know’

KNOW=ACT.PRS=1SG NEG=1SG KNOW=ACT.PRS

Otherwise, the position of the members of this group is fixed. See Table 3.1 above for a list of all the members of this category except for the markers =*kiyar* ‘DU’ and =*ki* ‘PL’

Group III: High mobility

Once again, the floating clitics form a separate group and, as their name suggests, are entirely free in their distribution as long as they do not directly precede another enclitic or follow a demonstrative. For examples, see 3.1.1.

3.2.6 *Lexicalization*

Lexicalization of units marked by a clitic is highly marginal in Kharia. The only likely candidates in the data are the following three units:

47. *diyoga* ‘daily’ from *diyo* ‘daily’ and =*ga* ‘FOC’

48. *la?ko* ‘but’ from *la?* ‘then’ and =*ko* ‘CNTR’

49. *caheko* ‘instead’ from *cahe* ‘or’ and =*ko* ‘CNTR’

The form *diyo* is possible without =*ga* with the same meaning, although its use without the focal marker is rather uncommon. *la?* and *cahe*, on the other hand, are found without =*ko*, but only with different meanings, as shown above.

3.3 AFFIXES

In the lack of any information to the contrary, it is assumed here that all lexical items and postpositions which show the typical LH pitch pattern are also grammatical words (or “syntactic atoms”). This is in line with the fact that enclitics may attach to these elements.

It is further assumed here that ANY units that clitics can attach to are grammatical words. This will allow us to consider the v2s, whose status as phonological words is uncertain (2.5), to be grammatical words as well, since in the presence of more than one v2, non-final v2s may also be marked by a floating clitic:

50. *ka?bto sango?d dom=ga may=ki.*
 door close v2:PASS=FOC v2:TOTAL=MID.PST
 ‘The door was shut entirely.’

This is in contrast to those units which are neither potential phonological words nor enclitics, i.e., those elements which can uncontroversially be considered affixes. These units must always combine with some other unit and are always integrated into the LH-pattern.

There are relatively few affixes in Kharia. The following presents a brief but largely exhaustive list of all elements in Kharia which may be considered derivational affixes.

51. Causative markers:

- *oʔb-* (prefix), *<ʔ>* / ** / *<ʔb>* (infix)—The distribution of these allomorphs is primarily determined according to the number of syllables of the underlying contentive morpheme but is partially idiosyncratic (cf. 6.3.5). Cf. also Diagram 2.14 in Chapter 2.
- *<a>*, *<(u)w>* (infixes, found only in contentive morphemes of Sadri origin ending in /e/) (cf. 6.3.3; 6.3.5).

Intensive markers: *-boʔ*, *-daʔ*, *-son* (all suffixes, lexically determined, cf. 5.10.2)

Participial marker: *-l* (suffix, found only in contentive morphemes of Sadri origin ending in *-ay*, cf. 6.6.2.5)

-duʔ: meaning unclear; found in “participial” usage (cf. 6.6.2.5) but also with a number of lexical morphemes to denote an inherently durative action (cf. *yo* ‘see’, *yo-duʔ* ‘stare at’), as well as in a few other forms (*konsel* ‘girl’, *konsel-duʔ* ‘woman’).⁶ With contentive morphemes which can denote a property, *-duʔ* often has an approximative meaning, “-ish”, and will be glossed as such: *maha-duʔ* [big-APPROX] ‘biggish’ Its exact status requires further research.

-nV- traditionally considered a “nominalizing” infix, in which *-V-* has the same quality as the vowel of the underived contentive morpheme. Not productive: *joʔ* ‘sweep’, *jo<no>ʔ* ‘broom’ Cf. 4.3.3.

There do not seem to be any other derivational affixes in Kharia, with all other grammatical marking being expressed by grammatical units which are either phonological words (i.e., the reciprocal marker *kol*, discussed in 6.3.5), through the enclitics discussed in 2.5 and this chapter, or units whose status is at present indeterminate between these two groups (the so-called “v2s”, discussed in 6.5).

⁶ It is possible that *-duʔ* is related to the Juang “article” *-de* (Matson, 1964: 44), e.g., **kon* ‘small’ / *konduʔ* ‘child’ (≈ ‘the small one’), although if so the Kharia-internal development of some contrasting pairs remains uncertain: *konsel* ‘girl’ vs. *konselduʔ* ‘woman’, *yo* ‘see’ vs. *yoḍuʔ* ‘stare at’ or *rusuŋ* ‘red’ vs. *rusuŋduʔ* ‘reddish’.

PARTS OF SPEECH AND THE LEXICON

4.1 INTRODUCTION AND METHODOLOGY

In Kharia, or rather, in the Simdega dialect of Kharia (cf. 4.2 below), what correspond to noun phrases and verbs in English have the same underlying structure. For reasons which will become clearer in the following, these will be referred to here according to their structure as the “Case-syntagma” and the “TAM/PERSON-syntagma”

As will be shown in the following pages, TAM/PERSON-syntagmas and Case-syntagmas in the dialect of Kharia spoken in and around the city of Simdega have the same underlying structure, a structure which is formed at the syntactic (i.e., “phrasal”) level, not in the lexicon. It is the functional head of this unit which signals the status of this constituent: With Case-syntagmas this is “case”, taken here to include adpositions,¹ with TAM/PERSON-syntagmas it is TAM/basic voice (active / middle) and person / number / honorific marking, all of which are enclitic. The only restriction which has come to our attention is that CASE may not combine with TAM/BASIC VOICE and PERS/NUM/HON-marking. Otherwise, semantic heads (or “bases”)² which can denote physical objects or properties may appear in predicative function as well as attributive and referential function, while semantic heads which can refer to actions may appear in referential and attributive function, in addition to predicative function.

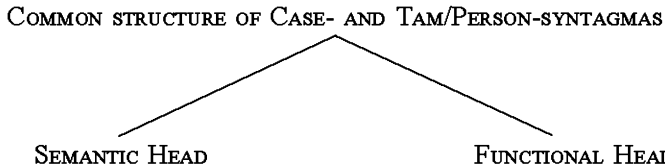


Diagram 4.1: Common structure of TAM/PERSON- and Case-syntagmas in Kharia

¹ Although excluding the genitive, see Section 4.4 for further discussion.

² The terms “semantic head” and “semantic base” will be used interchangeably throughout this study, with no difference between the two being implied.

The situation in the Gumla dialect, as we shall see below, is different in this respect. Although the general structure of TAM/PERSON-syntagmas and Case-syntagmas is the same as given above, speakers of this dialect in interviews did not accept “phrasal” semantic heads in TAM/PERSON-syntagmas but rather only semantic heads consisting of one or two contentive morphemes, neither of which may refer to a physical object.

In the remainder of this section, the methodology used in this study to determine parts of speech such as noun, verb and adjective will be discussed, focusing almost entirely on structural properties, before proceeding in the remaining sections of this chapter to discuss the data in detail.

The individual parts of speech are defined here as follows, based largely on the definition found in Hengeveld (1992: 37) and Hengeveld & Rijkhoff (2005: 406f.), although differing slightly from this.

- VERB: a lexeme that without further measures being taken can be used as the head of a predicate phrase ONLY
- NOUN: a lexeme that without further measures being taken can be used as the head of a referential phrase ONLY
- ADJECTIVE: a lexeme that without further measures being taken can be used as a modifier of the head of a referential phrase ONLY
- (MANNER) ADVERB: a lexeme that without further measures being taken can be used as a modifier of a non-nominal head ONLY

To begin with, note that all four categories refer to LEXEMES. As parts of speech in the traditional sense are lexical classes, this seems to be unproblematic, although it will turn out to be of importance in Section 4.4.

Following Hengeveld (1992) and Hengeveld & Rijkhoff (2005), a verb is defined here as a lexeme which is restricted to the head of a predicate phrase. If a verbal lexeme appears in any other function, then “further measures” such as derivational marking must be taken. Unlike Hengeveld (1992) and Hengeveld & Rijkhoff (2005), however, this same criterion is applied here to nouns, adjectives and adverbs. That is, these categories are restricted to the uses given above IF NO FURTHER MEASURES ARE TAKEN, whereas in both of the studies just mentioned, it is assumed that all lexemes may function as “verbs”. The reason we differentiate here somewhat more stringently is that “non-verbs”, when functioning as the predicate, typically require some type of overt “verbalizing” marking, whether derivational, as with *diverse* → *diversify*, or morphosyntactic, e.g., a copula verb, as in *be happy*. Hence, in the present study non-verbs are restricted to the respective functions given above if no further measures are taken.

Following Sasse (1993: 96ff.), it is also assumed here that parts of speech in general must be analyzed from at least the following four perspectives:

- *Formal parameters*—this includes inflection, derivation and the possible distribution of the elements of the lexicon (i.e., can they co-occur with demonstratives, with predicative marking, with case marking, etc.);
- *Syntactic parameters*—These may be of two types:
 - Phrasally defined configurational categories—NP, VP, AP, etc.
 - Functionally defined—predicate, argument.
- *Ontological-semantic parameters*—membership in ontological categories or semantic classes
- *Discourse-pragmatic parameters*—these refer to the basic discourse operations such as reference, predication and modification.

In order to show that a language has a noun / verb / adjective distinction in this perspective, we must be able to demonstrate the following:

- The lexicon can be divided into lexical classes according to structural (= formal) properties (e.g., do certain lexemes require derivational marking to appear in certain functions?);
- If lexical classes of some type can be found, do these classes correspond directly to a syntactic category?
- If so, do these syntactic classes correspond at least primarily to discourse-pragmatic parameters, such as reference vs. predication?
- Only if we have been able to show that the above three criteria hold should we consider the ontological properties of the individual lexical items. If it can be shown that the lexical classes correspond directly to syntactic functions which correspond to reference-discourse functions, and that this correspondence depends at least “prototypically” on certain ontological classes (such as physical objects, events/actions, etc.) will we be justified in assuming that the language in question has nouns, verbs, and adjectives as lexical classes.

This of course presumes that predicates and (potentially) referential units have LEXICAL heads. As we shall see in the following, however, this assumption is not borne out in Kharia. Thus whether or not we find all “lexemes”—or in our terminology, contentive morphemes—in Kharia to

be “pre-categorial”, the importance of this criterion for Simdega Kharia will be seriously called into question.

“TAM/PERSON-syntagma” and “Case-syntagma” are purely STRUCTURAL categories: Although they are intimately linked with predication and reference, respectively, they cannot be equated entirely with these. First, TAM/PERSON-syntagmas may also contain referential morphemes. Furthermore, both TAM/PERSON-syntagmas and Case-syntagmas in Kharia may also be used as modifiers, as in the following examples:

1. *u goʒa duniya=te lebu=ki=yaʔ kahani*
 this entire world=OBL person=PL=GEN story
 ‘the story of the *people* on this entire world’

2. *yo=yoʔj lebu col=ki.*
 see=ACT.PST.1SG man go=MID.PST
 ‘The man I saw left’ (= literally: ‘I=saw man left’.)

TAM/PERSON-syntagmas may also be used referentially, (3) and Case-syntagmas predicatively (4).

3. *kunqab aw=ki tomlin kharīya gam qom=na laʔ=ki=may*
 behind QUAL=MID.PST milk Kharia say PASS=INF IPFV=MID.PST=3PL
ina no u=ki tomlin uʔd=ga qel=ki=may.
 because this=PL milk drink=FOC come=MID.PST=3PL
 ‘[Those who] were in the rear were called “Milk Kharia” because they came drinking milk.’ [MT, 1:180]

4. *ro u=ga ho jinis=aʔ komaŋ.*
 and this=FOC that animal=GEN meat
 and this [is] *that animal’s meat.*” [AK 1:57]

One last aspect of the determination of parts of speech that must be discussed here is that of semantics. Unlike other authors such as Evans & Osada (2005a) or Croft (1991; 2000; 2005), no great emphasis is placed here on the semantic predictability of the lexical items in various functions, although as will be seen in the following pages, semantic predictability in Kharia is certainly very close to 100%. The reason for this is that even in languages with well-defined word classes such as English, one and the same lexical item can have quite different meanings in different

contexts even though it appears in these different contexts in the same part of speech. Cf. e.g., *bank* in English, which as Pustejovsky (1998: 28) notes can have the meaning of a financial institution (*The bank raised its interest rates yesterday*) or the construction in which this financial institution operates (*The store is next to the newly constructed bank*). Similarly, *newspaper* (the physical object or the producers of this physical object), etc. Clearly, if we find this type of variation for a single lexical item within a single word-class category, it would be unreasonable to expect that it should be ENTIRELY absent when the same contentive morpheme is used predicatively or attributively in languages where this type of use is possible. The only requirement is that the various meanings of this lexical item are related to one another in a systematic way.³

The remainder of this chapter is structured as follows. We begin in 4.2 with a traditional discussion of “precategoriality” in Kharia where all contentive morphemes may appear in predicative, referential or attributive function, given the appropriate context. Although this section is not primarily concerned with semantics, the various lexical morphemes are divided loosely into “ontological classes” as a convenient means of presenting these.

After having dealt with underived contentive morphemes in some detail, 4.3 then examines three derivational morphemes/processes involving infixation and reduplication which are often considered nominalizers in Munda languages and which require some discussion.

In 4.4 it will be shown that what at first glance appear to be “nouns” and “verbs” in Kharia are in fact not lexical categories at all but rather syntactic constructions, neither of which can be equated with the endocentric categories NP/DP or VP/IP. Section 4.5 then presents a summary of the discussion of parts of speech in Kharia.

Finally, Section 4.6 deals with other issues related to the lexicon, primarily the so-called “compounds” and “nominal incorporation”, while Section 4.7 deals with the “echo-word” formation.

4.2 THE “PRECATEGORIALITY” OF SIMPLE CONTENTIVE MORPHEMES

In the present study the term “precategoriality” is used to mean that any contentive morpheme can be used referentially, predicatively, and in

³ See also Hengeveld & Rijkhoff (2005: 415f.) for a similar view on this issue.

following discussion, as in the rest of this grammar, concentrates on the Simdega dialect, as these are the speakers we worked with the most. Where dialectal differences became apparent, however, these will be referred to in the text itself and this information will be summarized in 4.4.1 and 4.5. When no reference is made to the dialect being referred to, the Simdega dialect is meant.

The (Simdega) Kharia lexicon may be divided into two broad groups—one open group and a much smaller, closed group, which sub-divides into two further closed groups. The large open class consists of the contentive morphemes, all of which can be used predicatively, referentially or in attributive function. This class includes all morphemes which are potentially referential, such as *table*, *notebook*, or *dog*, as well as events and states such as *run*, *eat* and *beautiful*. This holds true of all contentive morphemes, whether of Kharia origin or loan words.

The other group, the closed group, is much smaller and consists of two further sub-classes. The first sub-class are the functional or “grammatical” morphemes, which can never be used as a predicate, as an attribute, nor in referential function. This group includes markers for tense, aspect, mood, person marking, case (including postpositions), etc., i.e., those elements which provide the framework or grounding of the clause within a particular setting (e.g., TAM) or which integrate the various constituents into the clause by indicating the function of a constituent, e.g., case.

The second sub-class consists of proforms and temporal and locative deictic elements. These may freely appear in referential or predicative function with no derivational marking, “light verbs”, etc., but appear to require the genitive when used attributively. This is illustrated in Diagram 4.2.

The following demonstrates the functional flexibility for a number of lexical items from the open class and the small class of proforms / deictics, divided here for convenience into various ontological classes. No attempt at exhaustivity is made here. For numerous further examples, consider the entries in Peterson (2009).

Interrogatives

Interrogatives, which generally appear in referential function, can also appear in attributive function or as the semantic head of a predicate with no derivational marking. In the latter case, the middle voice denotes a change of state, *i=na?* [what=MID.IRR] ‘what will happen / become?’,

as a convenient way of referring to these two varieties, which otherwise are highly similar to one another.

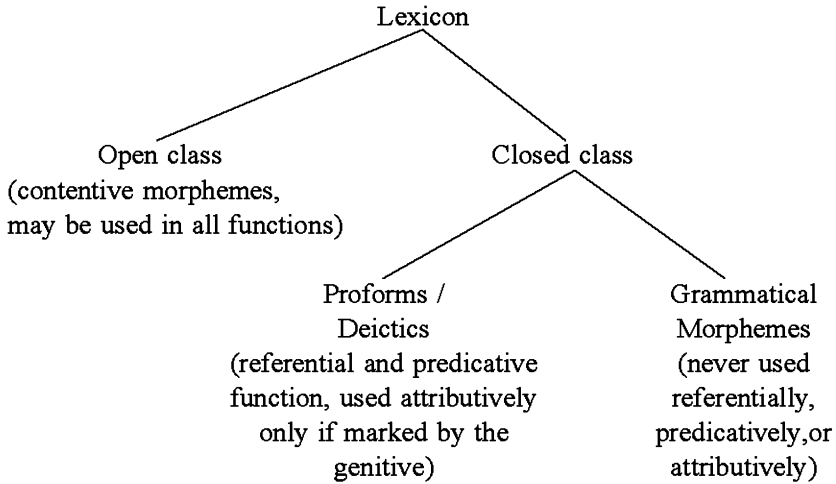


Diagram 4.2: The Kharia lexicon

whereas the active denotes an activity, *i=ye=m?* [what=ACT.IRR=2SG] ‘What will you do?’ The following presents a few examples.

Referential function

6. *i karay=oʔb?*
 what do=ACT.PST.2SG
 ‘What did you do?’

Predicative function

7. *gam=e=pe i=ye=niŋ?*
 say=ACT.IRR=2PL what=ACT.IRR=1PL.INCL
 ‘Tell [me], what shall we do?’ [Kerkeṭṭā, 1990: 6]

Attributive function

8. *aba=ḍom=kiyar um=kiyar batay=o? no i jinis*
 father=3POSS=HON NEG=HON tell=ACT.PST Cmpl what animal
heke, i jhāut heke lekin, muda, ḍaŋ goṭh=o?
 QUAL.PRS what animal QUAL.PRS but but send C:TEL=ACT.PST

sou²b=te=ga.

all=OBL=FOC

'Their father didn't tell them *which* animal it is, *which* animal it is, but he sent them all off.'

[AK, 1:10]

Indefinites

Similarly, indefinites can appear in any function without any overt derivational morphology:

Attributive function

9. *babu, kimin=ki jahã jio?=na cij dho?=si?=pe*
 child daughter.in.law=PL INDEF.NHUM eat=INF thing grab=PERF=2PL

gur=na,

fall=MD.IRR

'Child, *some* of the food (= *some* to=eat things) which you daughters-in-law have taken will fall,

[MS, 1:44]

"Referential" function

10. *la? am na cori karay=na, na jahãy=a? jan*
 then 2SG NEG theft do=INF NEG INDEF.HUM=GEN life

tar=na, na jahãy=a? jahã karay=na.

kill=INF NEG INDEF.HUM=GEN INDEF.NHUM do=INF

'So do not steal, do not kill anyone, do not harm anyone (= do not do anyone's *anything*).'

[MS, 1:309]

Predicative function

11. *jahã guɽu?=may,⁵ aniŋ=te sa?dhe=si?=may, ho=je?*
 INDEF OPT=3PL 1PL.INCL=OBL torment=PERF=3PL that=SG.NHUM

bhagwan=jo=ko yo=te?²j.

God=ADD=CNTR see=ACT.PROG

⁵ *guɽu?* 'OPT' is not a verbalizer and is compatible with any predicate which can be agentive. Cf., e.g., *col guɽu?* 's/he should go'

‘Let them do as they like (= *let them [do] whatever*), they have tormented us, and God is watching that as well.’ [Kerkeṭṭā, 1990: 7]

Quantifiers

In addition to their usual modifying function, quantifiers may also serve as predicates and be used referentially. The meaning of the quantifier in predicative function is that of ‘become X’ (MIDDLE) and ‘turn into X’ (ACTIVE).

Attributive function

12. *la?* *beṭ=ḍom=ki* *raji=ki=may* *ro* *kinir* *col* *god=ki.*
 then son=3POSS=PL accept=MID.PST=3PL and forest go C:TEL=MID.PST

la? *moṇ* *harin* *tar=o?*=*ki.*
 then one deer kill=ACT.PST=PL

‘Then his sons agreed and they went off to the forest. Then they killed a (= *one*) deer.’ [MS, 2:10]

Referential function

13. *ho=kaṭ* *muḍu=te* *juṇ=o?* ”⁶
 that=SG.HUM one.HUM=OBL ask=ACT.PST
 ‘He asked *one* [person].’ [HJPa:67]

Predicative function

14. *jhaṛi* *daru* *biru* *oḍo?* *moṇ* *god=ki=may.*⁷
 all tree mountain other one C:TEL=MID.PST=PL
 ‘All trees and mountains and everything else (= other) *became one.*’ [HJPa:118]

Similarly: *thorek* ‘few, a little; decrease (ITR/MIDDLE; TR/ACTIVE)’

⁶ Note that *muḍu* refers only to humans whereas *moṇ* refers only to non-humans. Otherwise, their meaning is identical, i.e., ‘one’

⁷ Note that *god=ki* ‘C: TEL’ is a so-called “v2” (6.5.3), not a verbalizer. It is e.g. also found with actional predicates, cf., e.g., *col god=ki[=may]* [go C:TEL=MID.PST=3PL] ‘they went’ in 12. above.

ubar 'two; become two (MIDDLE); make two (ACTIVE)' and many, many others.

Properties

Any typically attributive contentive morpheme can also be found in predicative function.

Attributive function

15. *ho* *rusuŋ* *o?* 'that *red* house'
 that red house

Predicative function

16. *am=bar* *rusuŋ* *go?d=ki=bar!*
 2=2HON red C:TEL=MID.PST=2HON
 'You've *gotten all red* (= i.e., you have a sunburn)!'

In referential function these may denote either the abstract notion, such as 'redness', or reference to an object, i.e., 'red one' Unfortunately, the data are not sufficient to demonstrate this for each entry in this group, however, the following examples for the contentive morpheme *ranga* 'cold (adj./n.)' demonstrate the underlying principles:

17. ... *ijn=te* *ranga* *bhere=ya?* *gha²d* *terom* *thuray=na*
 1SG=OBL cold time=GEN for honey gather=INF
 ayi²j.
 QUAL.PRS
 'I have to gather honey for the *cold* season (= time).' [TK, 2:20]

18. *ranga=ya?* *dino* *d²am=ki.*
 cold=GEN day arrive=MID.PST
 'A cold day (= day *of coldness*) arrived.' [TK, 2:28]

19. *kosu* *ranga* *buŋ* *bhiren=ta²d=ijn.*
 sickness coldness INST flounder=MID.PROG=1SG
 'I am plagued by (= floundering with) sickness and illness (= *coldness*).' [AT, 4:2]

In predicative function *ranġa* has the meaning ‘feel cold (MIDDLE); make cold (ACTIVE)’

Similarly, in reference to an object (cf. English ‘the big one’) consider *maha=ki* in (20). Note also the predicative use of *konon* ‘(become) small’ here with the slightly different meaning ‘become impure’. This also shows that such semantic shifts do not require us to posit a new contentive morpheme or word category, especially since in this case *konon* is attested elsewhere in the data with the expected meaning ‘become small’, as in (21).

20. *maha=ki goʔjhuŋ hinte i no i jait=yaʔ*
 big=PL path LOC what or what ethnic.group=GEN
lebu=ki=yaʔ peʔ jokh=oʔ=ki ro konon goʔdʒ=ki=may.
 person=PL=GEN rice eat=ACT.PST=PL and small C:TEL=MID.PST=3PL
 ‘The elder ones (= *big [ones]*) ate the food (= rice) of people of various tribes along the way and [thus] *became small* [i.e., impure].’
 [MT, 1:27]

The following example was said to this author jokingly by a friend before entering a person’s home after this author had repeatedly hit his head in Simdeġa on low doors while entering cars, buses or people’s homes:

21. *konon=na=bar!*
 small=MID.IRR=2HON
 ‘Become small! / Make yourself small!’

The following presents a few of the many contentive morphemes belonging to this class:

- ebloŋ* ‘alone; being alone; become alone (MIDDLE); cause s.o. to be alone (ACTIVE)’
hoʔbne ‘enough (n./adj.); become enough (MIDDLE); make s.th. enough (ACTIVE)’
maha ‘big (one); very; grow, become big (MIDDLE); enlarge (ACTIVE)’
ranġa ‘cold; coldness; feel cold (MIDDLE); make cold (ACTIVE)’

Proper names

Proper names may also function as the semantic head of both TAM/PERSON-syntagmas and Case-syntagmas, with the by now familiar semantic interpretations.

22. *a?ghrom*

‘Aghrom’ (name of a town) [e.g., as a subject, object, goal (adjunct), etc.]

23. *a?ghrom=ki*.

Aghrom=MID.PST

‘It came to be called (= became) Aghrom.’

24. *a?ghrom=o?*.

Aghrom=ACT.PST

‘[S/he] named [it] (= turned [the city] into) Aghrom.’

We have no example of *a?ghrom* in attributive function in our data, but consider the following example of *rāci* ‘Ranchi’ (capitol of Jharkhand state) in attributive and referential function:

25. “*ani yar=te=niŋ!*” *gam=o?=ki ro ho=ki*
 come.on! flee=ACT.PRS=1PL.INCL say=ACT.PST=PL and that=PL
- qel=ki=may u=ti?j choṭanagpur pahaṭi prades ti?j*
 come=MID.PST=3PL this=side Chotanagpur mountain region side
- u=ti?j qel=ki=may rāci eriya=te rāci mane*
 this=side come=MID.PST=3PL Ranchi area=OBL Ranchi umh
- qel=ki=may.*
 come=MID.PST=3PL
 ‘They said “Come on, let’s go!” and they came this way, towards the mountainous Chotanagpur region, they came this way, to the *Ranchi area*, to *Ranchi*, umh, they came.’ [MS, 1:211]

Status and Role

Terms referring to the status of people or their role can also occur in the different functions:

Referential function

26. *u kayom onḍor=kon raṭa=ya? ayo=ḍom, darhi=ya?*
 this matter hear=SEQ Rata=GEN mother=3POSS Darhi=GEN
saw-ṭay=ḍom, gam=te:
 spouse-woman=3POSS say=ACT.PRS
 ‘Hearing this matter, *Rata’s mother*, [i.e.,] Darhi’s wife, said:
 [BB, 2:66]

Predicative function

27. *ho=kaṭ ayo=ki.*
 that=SG.HUM mother=MD.PST
 ‘She *became a mother.*’
28. *ho=kaṭ=te ayo=yoʔj.*
 that=SG.HUM=OBL mother=ACT.PST.1SG
 ‘I *accepted her as [my] mother.*’

Note that the use of *ayo* in the active differs only very slightly from the expected meaning ‘turn [s.o.] into a mother’ We consider this entirely predictable since ‘turning someone into a [i.e., one’s own!] mother’ involves a change in attitude on the part of the subject as well as in his or her behaviour towards the new mother, i.e., ACCEPTING someone as a mother.

Similarly: *apa* ‘father; become a father (MIDDLE); accept s.o. as a father (ACTIVE)’
jait ‘ethnic group; become a Kharia (MIDDLE); accept s.o. as a Kharia (ACTIVE)’⁸

On the use of these morphemes in attributive function, see the comments on “compounds” in the following.

⁸ Obviously, ‘ethnic group’ in the predicative function is at least primarily taken as referring to one’s own ethnic group, i.e., the Kharia, although I have no doubt that it could also refer to other ethnic groups in a complex predicate of the type discussed in 4.4 below.

Physical objects and animate entities

A few examples of contentive morphemes are merely presented here for this category, as the semantics of this group are entirely predictable:

29. *cakħna?* 'curry; become a curry (MIDDLE); make s.th. into a curry (ACTIVE)'
kaḍoŋ 'fish (n.); become a fish (MIDDLE); make s.th. a fish (ACTIVE)'
lebu 'man; person; become man (MIDDLE)' (cf. example (5))
lutur 'ear; become an ear (MIDDLE); make s.th. an ear (ACTIVE)'

Contentive morphemes such as these are also commonly found in attributive function, where they are traditionally assumed to form compounds with the following contentive morpheme. However, in Section 4.6 it is argued that these are in fact not compounds at all but merely two (or more) juxtaposed contentive morphemes which do not form a tighter syntactic or phonological unit.

It is worth noting in this respect that this flexibility also holds for loan words, including recent loan words, cf. *tebul / tebal* 'table (n.); table-; become a table (MIDDLE); turn s.th. into a table (ACTIVE), from English *table*. This is an entirely productive pattern.

Locatives

Lexical items denoting a location can also appear in the various functions. In predicative function, these contentive morphemes generally appear in both the middle (ITR) and active (TR) voices.

30. *toblun* 'top; on; go up, rise (MIDDLE); take up, raise (ACTIVE)'
tuta 'down; bottom; below; go down (MIDDLE); put down, lower (ACTIVE)'

Activities

Contentive morphemes denoting activities can also occur in predicative, referential or attributive function. It should be noted that the infinitive is preferred in referring to an action, nonetheless the following examples were all accepted as grammatical.

karay ‘do’

Predicative function

31. *lekin am pujapaṭh karay=teʹjd=em, ...*

but 2SG sacrifice do=ACT.PROG=2SG

‘But you *are doing* the sacrifice,

[AK, 2:32]

Attributive function

32. *biha karay lebu goʹj goʹd=ki.*

marriage do man die C:TEL=MID.PST

‘The man who married (= *did* marriage man) died.’

Referential function

33. *u=ghay karay um=ijn baʹj=ta.*

this=way do NEG=1SG like=MID.PRS

‘I don’t like *doing* [it] like this.’

siloʹ ‘plow’

Predicative function

34. *modi na siloʹ=na laʹ=ki no oḍoʹ=ga hoḍom kamu*
Modi neither plow=INF IPFV=MID.PST or more=FOC other work

karay=na laʹ=ki.

do=INF IPFV=MID.PST

‘Modi neither *plowed* nor did any other work.’

[RD, 2:39]

Referential function

35. *ho=kaṛ=aʹ siloʹ yo=yoʹj.*

that=SG.HUM=GEN plow see=ACT.PST.1SG

‘I saw his *plowing*.’

kayom ‘talk’

Predicative function

36. *kahani lebu=ki khor̥i=ki=te kayom=ta=ki u=ghay*
 story person=PL village.section=PL=OBL speak=MID.PRS=PL this=way
ayiʔj.
 QUAL.PRS

‘The people *tell* [this] story in the villages. It goes like this.’ [AK, 1:2]

Referential function

37. *u kayom on̄dor=kon raʔa=yaʔ ayo=d̄om, dar̄hi=yaʔ*
 this speak hear=SEQ Rata=GEN mother=3POSS Darhi=GEN
saw-ʔay=d̄om, gam=te.
 spouse-woman=3POSS say=ACT.PRS

‘Hearing this matter (= *talk*), Rata’s mother, Darhi’s wife, says: “[. .]”
 [BB, 2:66]

There is a slight complication to the distribution given above in that when non-borrowed contentive morphemes are found anywhere within a Case-syntagma, the form must be polysyllabic. If the lexical item is monosyllabic, it is obligatorily reduplicated, as the following examples show.⁹

Predicative function

38. *id̄ib=te moʔʔho moʔʔho m̄oloy ʔibru kuda kolon̄*
 night=OBL fat REP five six millet bread
ter=na laʔ=ki=may.
 give=INF IPFV=MID.PST=3PL

‘At night, *they used to give* them five or six big fat millet breads.’

[BB, 1:11]

⁹ This differs from what I refer to as “repetition”, as with *moʔʔho moʔʔho* in 38., which is facultative and can be used to denote intensity, plurality or extended temporal duration. “Reduplication” on the other hand is an obligatory, phonologically motivated process (cf. 4.3).

Attributive function

39. *burhikutain=jo* *u=ki=ya?* *ter-ter* *jimi=ga* *heke.*
 Burhikutain=ADD 3=PL=GEN give-RDP name=FOC QUAL.PRS
 ‘Burhikutain is also the name *they have given it.*’ [Kerkeṭṭā, 1990: ii]

Referential function

40. *ho=kaṭ=a?* *ter-ter* *yo=yoʔj.*
 that=SG.HUM=GEN give-RDP see=ACT.PST.1SG
 ‘I saw *his giving [it].*’

This issue will be discussed in Section 4.3.2 in more detail. For the moment it will suffice to note that Kharia has a strong bisyllabic constraint which requires phonological words to be at least bisyllabic. As the direct case is not overtly marked, or “zero marked” (5.2), this means that any element appearing in a Case-syntagma could potentially appear without case marking and would then be monosyllabic if the contentive morpheme is monosyllabic. In these cases, the contentive morpheme is obligatorily reduplicated. This does not hold, however, for the TAM/PERSON-syntagma, as person marking and/or TAM/basic voice marking (active/middle) is always present. Thus, the lexical head of a Case-syntagma must be at least bisyllabic, whereas that of the TAM/PERSON-syntagma need not.

Deictic “adverbials” and proforms

Temporal and locative deictics corresponding to English *here, there, today, tomorrow*, etc., normally appear in Kharia as adjuncts to the clause:

41. *iḍa?* *col=ki.*
 yesterday go=MID.PST
 ‘He went *yesterday.*’ [Kerkeṭṭā, 1990: 15]

When used as adjuncts, deictics are analyzed here as Case-syntagmas and not as a separate class of “adverbials” since they may also be followed by a marker of the Case-syntagma, such as, e.g., a postposition, *musa tay* [today ABL] ‘as of / beginning with today’ Also, it should be noted here that deictics, although a small, closed class, behave like the contentive morphemes in that they may combine with the genitive (42) and with

number marking,¹⁰ such as the “approximative” function of the plural in (43) (cf. 5.4).

42. *idg=a?* ‘yesterday’s’ (e.g., yesterday’s newspaper, special genitive base; cf. comments below)

43. *id̩a?*=ki *q̩el*=ki *hoy*.
 yesterday=PL come=MID.PST INFER
 ‘He came *yesterday or so*, I guess.’

With respect to predicative function: Although it is rather difficult to construct a situation where *id̩a?* can occur as a predicate, speakers did accept examples such as the following, provided that God is responsible for today’s becoming yesterday.

44. a. *musa id̩a?*=ki. b. *ponomosor musa=te id̩a?*=yo?
 today yesterday=MID.PST God today=OBL yesterday=ACT.PST
 ‘Today *became yesterday*.’ ‘God *made today yesterday*.’

What translate into English as pronouns may also appear in both referential and predicative function, as the following elicited example shows.

(in a play about me and you, in which both of us will be taking part):

45. “*naʔak=te iʔn=ga ho=kaʔ=na=iʔn ro am=ga*
 play=OBL 1SG=FOC that=SG.HUM=MID.IRR=1SG and 2SG=FOC
iʔn=na=m.” “*umbo?*” *am=na* *um=iʔn* *pal=te*.
 1SG=MID.IRR=2SG no 2SG=INF NEG=1SG be.able=ACT.PRS
q̩irekt̩ar seʔn=ga? *iʔn=te* *ho=kaʔ=o?* *am=ga*
 director early=FOC 1SG=OBL that=SG.HUM=ACT.PST 2SG=FOC

¹⁰ It will be shown in 4.4 that the genitive and number and possessive markers are not “nominal” or restricted to the Case-syntagma but may surface in both Case- and TAM/PERSON-syntagmas. Cf. Section 4.4 for further discussion. The purpose of these examples here is simply to show that deictics and proforms behave exactly like the contentive morphemes, with the exception of the obligatory genitive marking in attributive function. Although few would be surprised to find proforms marked for the genitive or number (cf. 5.6), deictics such as ‘today’ and ‘tomorrow’ might not be expected with this marking, hence these examples.

am=na=m.”

2SG=MD.IRR=2SG

‘“In the play I *will be him* and you *will be me*.” “No. I *can't be you*. The director already *made me him*. You *will be you*.”’

The reason for grouping deictics and proforms in the same category is purely structural: What these two categories appear to share—although this may be due to lack of data at the moment—is that they must appear in the genitive in attributive function. Or, put differently, the only data currently available for proforms or temporal / locative deictics appearing in attributive function is in the genitive, e.g. *idg=a? khabar* [yesterday=GEN news] ‘yesterday’s news’ and *in=a? gotar* [1SG=GEN clan] ‘my clan’. Thus, this group appears to form a subclass of the closed class and is restricted to only a few morphemes. This is also in line with the fact that many of these forms, especially the deictic morphemes, have special genitive forms, suggesting that they do in fact constitute a separate class in the lexicon, although this class does not seem to fit in well with an analysis of “nouns”, “adjectives” or “verbs”¹¹

This analysis is, however, somewhat uncertain: Note that the use of the genitive is more or less obligatory in Kharia when the “possessor” is [+definite], which is certainly the case with these forms (cf. Section 5.2). As such, the use of the genitive here may be less due to “lexical classes” than to definiteness. This issue requires further research. We will consider them a separate lexical class.

Summarizing, the few examples in the preceding sections are typical for hundreds of other contentive morphemes in that they may be used in referential, attributive and predicative function, and that the different meanings of the individual lexical items in different functions are highly predictable. There is also a small, closed class of proforms and deictics. These differ from the open class in that they may not appear in attributive function without genitive marking, although there are no restrictions on their use in referential and predicative function. Their meanings in different functions are as equally predictable as those of the open class.

Finally, there are a few contentive morphemes which would seem to deviate slightly from the expected meanings in predicative or referential function, although these deviations are so slight that it would seem unreasonable to consider them different lexical items. The following list is fairly exhaustive for such shifts in our data.

¹¹ Furthermore, the fact that monosyllabic members of this class never reduplicate, such as *in* ‘1SG’ and *am* ‘2SG’, also sets them off from the large, open class. Cf. the discussion in Section 4.3.2.

46. *deŋa* 'shaman; practice shaman rites (ACTIVE)'
haqa 'urine; urinate (ACTIVE)'
iʔjthan 'cow dung; defecate (of cows) (ACTIVE)'
juʔd 'root (of a plant); sprout roots (MIDDLE); take root firmly (ACTIVE)'
kaluwa 'dinner; eat dinner (ACTIVE)'
kiriya 'oath; make an oath (ACTIVE)'
khabar 'news; inform (ACTIVE)'
kharom 'ashes; pour ashes into something (ACTIVE)' (≈ 'turn into ashes'? 'make ashy?')
loŋoy 'shade; take shelter in the shade (MIDDLE); make shade (ACTIVE)'
nasta 'breakfast; eat breakfast (ACTIVE)'
ongher 'male servant; work as a male servant (MIDDLE); keep a male servant (ACTIVE)'
sangoʔd 'friend; become friends (MIDDLE); accompany someone (ACTIVE)'
suʔtrom 'thread (n.); thread (a needle) (ACTIVE)'
updes 'sermon; give a sermon (ACTIVE)'

It should first be noted that most of these morphemes and their meaning were obtained in interviews and most stem from one and the same speaker, hence it is not unreasonable to assume that at least some of them will turn out to have the "expected" meanings upon closer examination. However, it should also be noted that virtually all of these contentive morphemes in predicative function refer to the FUNCTION that the entity denoted by this contentive morpheme performs: a shaman (*deŋa*) is someone who practices shaman rites, a male servant (*ongher*) is someone who does the work of a male servant, shade (*loŋoy*) is something to take shelter in during the midday heat, etc. Hence, there is a clear semantic relation between the two and hence no need to assume that we are dealing with two separate contentive morphemes.

On the other hand, a number of others pick up on the manner in which something comes into existence, e.g., *haqa* 'urine' and *iʔjthan* 'cow dung' are of course produced in the two actions that these contentive morphemes denote when used predicatively.¹² Again, there is no need assume that these are two separate contentive morphemes, as this is in fact the usual meaning of contentive morphemes denoting physical entities in predicative

¹² In the terms of Pustejovsky (1998), the two functions of the forms listed in 46. when used predicatively correspond to the TELIC and AGENTIVE quale, respectively.

function. The “problem” in this case is simply the English translation, not the Kharia semantics.

On the basis of this data, we conclude that (Simdega) Kharia does not possess lexical classes in the sense of nouns, verbs and adjectives. The only lexical classes it does possess, structurally speaking, are the open class, consisting of contentive morphemes, and a closed class, which subdivides into two groups: “functional” or “grammatical” morphemes (case, TAM, etc.) and proforms and locative and temporal deictics. Although there are thus three “classes” in this sense, they do not correspond in any neat fashion to “nouns”, “verbs”, and “adjectives”

Having dealt with underived lexical morphemes and their semantics in some detail, we now turn our attention to derived forms, including the reduplication mentioned above.

4.3 DERIVED FORMS

There are three derivational categories in Kharia which pertain directly to the discussion of parts of speech. The first category, which is referred to here as the *masdar*, involves the reduplication of monosyllabic native (i.e., non-borrowed) morphemes in certain environments. The remaining two categories both involve infixation: The first involves the infix *-NV-* only, while the second combines the *-NV-* infix with reduplication. These will now be dealt with briefly. Before beginning our discussion however, we must first discuss a basic principle which plays an important role in Kharia: the polysyllabicity of phonological words.

Anderson & Zide (2002) present evidence from a large number of Austro-Asiatic languages which strongly suggest that Proto-Munda had what they term the “Bimoraic Constraint”, which required nouns to contain at least two morae. They write: “A minimally bimoraic shape appears to have been obligatory for free-forms of nouns [as opposed to their compound forms, cf. Section 4.6, JP] in Proto-Munda and many of its daughter languages, which necessitated (and in Gta? continues to necessitate) the combining of the root element with some derivational process (in particular, prefixation, infixation, reduplication, compounding, or suffixation.” (Anderson & Zide, 2002: 55). In their definition, a mora is linked to any one of the following: 1. any vocalic nucleus, 2. a consonant in word-initial position before another consonant, resulting in a “sesquisyllabic” structure of the type “minor syllable—major syllable”, and 3. a post-vocalic [ʔ] in final position (in some languages).

The data they list in tables (3) and (4) of their study (Anderson & Zide, 2002: 57-59) clearly indicate that Kharia also once had this bimoraic

constraint for phonological words and that at least the following three means were employed to make a monomoraic form bimoraic: 1. The glottal stop, cf. Kharia *ti?* ‘hand’ with Juang *iti*, Gta? *titi*, etc. (but Santali *ti*); 2. reduplication, cf. Kharia *sajsaŋ* ‘turmeric’ vs. Gutob *sasaŋ*, Gta? *ssia*, and 3. the infix *-n-* (here referred to as *-NV-*, see below).¹³ Thus, assuming a syllable with the structure CV by way of example, deriving a bimoraic structure in Kharia would have resulted in one of the following structures: 1. CV?, 2. CV-CV, 3. CV-*n*-V

Finally, the authors note that each language subject to the bimoraic constraint seems to have favored one strategy over the others to fulfill this constraint (e.g., Anderson & Zide, 2002: 56). As we shall see in the following discussion, in Kharia reduplication has become virtually the only means of deriving bisyllabic phonological words, although there are remnants of other strategies, most notably the *-NV-* infix.

4.3.1 *Bisyllabicity in Kharia*

While it is assumed here that Anderson & Zide’s hypothesis is correct that a bimoraic constraint held in Proto-Munda, a subtle change has taken place in Kharia: The BIMORAIC constraint on free-forms in Proto-Munda has given way to a BISYLLABIC constraint on these forms in modern Kharia. The difference between the two in Kharia is only noticeable in one type of syllable: Syllables with the form CV? are monosyllabic but bimoraic according to the definition in Anderson & Zide (2002). Thus, in earlier stages of Kharia a bimoraic syllable like *ru?* ‘open’ would have counted as a free-form but in the modern language it must be made bisyllabic through reduplication, e.g., *ru?-ru?*.

Due to the structure of the language, this constraint has come to apply only to Case-syntagmas in Kharia: As the markers which signal a Case-syntagma, i.e., case, may be lacking (or a “zero morpheme”, cf. 5.2), a monosyllabic word could result, which should be avoided. This is different with a finite TAM/PERSON-syntagma, which is ALWAYS overtly marked for at least one grammatical category (TAM and/or person marking, cf. 6.1) and, with that, is always at least bisyllabic.

¹³ Another means referred to by the authors is the ‘potentially non-random patterning in Kharia of the prefix *kon-* in words referring to small animals or words roughly falling into a category of ‘diminutive’ or ‘affectionate’; ...’ (Anderson & Zide, 2002: 56). I believe that—as the authors’ quote implies—the use of *kon-* was restricted to cases referring to small entities. As such, although it could fulfill the bimoraic constraint, as it creates a second syllable and thereby adds one mora to the base, it was not primarily motivated by this constraint but rather simply compatible with it.

This principle has grammaticalized further in Kharia, so that in the modern language an underived, monosyllabic native contentive morpheme in the semantic base of a TAM/PERSON-syntagma need never be reduplicated. On the other hand, such forms in the semantic base of a Case-syntagma must always be reduplicated (with very few exceptions, to be discussed below), whether or not they carry further marking. This is thus a requirement of the syntactic construction as a whole, not of the morpheme itself.

We will now deal individually with the three derivational processes which serve to derive polysyllabic forms from underlying monosyllabic morphemes: reduplication, the *-nV-* infix and the combination of these two processes.

4.3.2 Reduplication—*The masdar*¹⁴

When appearing within a Case-syntagma, the semantic base is not an underived contentive morpheme but rather a masdar. This form may govern an object and adjuncts, which then take the same marking they would have with a finite TAM/PERSON-syntagma. What corresponds to the subject of the finite TAM/PERSON-syntagma appears in the genitive with the masdar. The masdar may also mark for case. Its functions will be dealt with in more detail in 6.6.2.1. Here it will only be dealt with in respect to its importance for a discussion of parts of speech in Kharia.

Note that the form of the masdar depends on the number of syllables AFTER any derivational processes, not of the underived contentive morpheme: Polysyllabic derived forms (e.g., underived contentive morpheme + causative or passive marker, if present) undergo no changes, while monosyllabic forms are reduplicated. This is thus a PURELY PHONOLOGICAL process.

47. <i>Underived contentive morpheme</i>	<i>Masdar</i>
<i>bay</i> ‘make’	<i>bay-bay</i>
Cf. <i>oʔb-bay</i> [CAUS-make]	<i>oʔb-bay</i>

¹⁴ In earlier studies on Kharia (e.g., Peterson 2007), the masdar is referred to as the “free-standing form”, as its form is ultimately due to the purely phonological constraint that free-forms be bisyllabic. However, as will be shown below, this form has grammaticalized further to become a grammatical category of its own. Although it is specialized for secondary predication, it is nevertheless also found in primary predication (Peterson & Maas, 2009). As it is not a “free-standing form” in primary predication, this term is inappropriate for this grammatical category and the more general term “masdar” will be used. See 6.6.2.1 for further details.

<i>borol</i> ‘live’	<i>borol</i>
<i>kerson</i> ‘marry’	<i>kerson</i>
<i>osel</i> ‘become white’	<i>osel</i>
<i>ru?</i> ‘open’	<i>ru?-ru?</i>
Cf. <i>ob-ru?</i> [CAUS-open]	<i>ob-ru?</i>
<i>rusuŋ</i> ‘become red’	<i>rusuŋ</i>
<i>soŋ</i> ‘buy’	<i>soŋ-soŋ</i>
Cf. <i>ob-soŋ</i> [CAUS-buy] ‘sell’	<i>ob-soŋ</i>
<i>yo</i> ‘see’	<i>yo-yo</i>
Cf. <i>oʔb-yo</i> [CAUS-see]	<i>oʔb-yo</i>

The masdar is perhaps most commonly found in what translates as a relative clause (48), but it may also be the semantic head of the Case-syntagma (49).

48. *iŋ=a?* *dura=te* *ru?-ru?* ***kunji***
 1SG=GEN door=OBL open-RDP key
 ‘**the key** I opened / open / will open / should open... the door with’

49. *ho=kaɽ=a?* *ter-ter* *yo=yoʔʔj.*
 that=SG.HUM=GEN give-RDP see=ACT.PST.1SG
 ‘I saw *him* give / *his* (act of) giving [something to someone].’

However the masdar can also be the semantic base of a primary predicate, which must then appear in the middle voice (6.4.2.2). Consider the following pair of examples demonstrating this use.

50. *iŋ* *ɖa?* *biʔh=oʔj.* *Simple contentive morpheme as semantic base*
 1SG water pour.out=ACT.PST.1SG
 ‘I poured water out.’

51. *iŋ* *ɖa?* *biʔɖ-biʔɖ=ki=iŋ.* *Masdar as semantic base*
 1SG water pour.out-RDP=MID.PST=1SG
 ‘I used to pour water out (e.g., that was my job, so I did it all the time)’

The morpheme *biʔɖ* ‘pour out’ in unmarked usage is restricted to the active voice, as in (50), and is underspecified with respect to habituality: ‘I poured out (once/twice/over and over/for years and years).’ The use of the reduplicated masdar as the semantic base, on the other hand, is only compatible with the middle voice and is explicitly marked for habituality, i.e., (51) cannot be interpreted as referring to a single event.

The use of the reduplicated masdar in TAM/PERSON-syntagmas in conjunction with the middle voice is thus morphologically and semantically marked and denotes habituality. With non-reduplicated polysyllabic morphemes, however, the use of this construction is not one of habituality but rather to express either a remote past tense or an indefinite, uncertain future event or situation.¹⁵

52. *goreʔj* *hakʔe=yoʔ.* ‘The ox grunted.’ (unmarked)
 ox grunt=ACT.PST

53. *goreʔj* *hakʔe=ki.* ‘The ox grunted.’ (a long time back)
 ox grunt=MID.PST

The use of non-reduplicated polysyllabic morphemes in this construction also serves as a further shibboleth for the two dialects discussed above in Section 4.2: Its use is restricted to the Simdega dialect, whereas speakers of the Gumla dialect consistently rejected its use in this construction. These facts, taken together, also show that the reduplicated masdar with its habitual interpretation is iconic to a certain extent,¹⁶ whereas the non-reduplicated masdar does not primarily have this habitual interpretation and is restricted to the Simdega dialect.

Although there is a semantic difference between a TAM/PERSON-syntagma with an underived contentive morpheme as its semantic base and one with a masdar as the semantic base (in conjunction with the middle voice), the fact remains that THE MASDAR IS NOT INHERENTLY “NOMINAL” NOR “ADJECTIVAL” BUT RATHER A “PRE-CATGORIAL” DERIVED FORM AND MAY BE USED REFERENTIALLY, ATTRIBUTIVELY OR PREDICATIVELY. It is for this reason that this unit is referred to here as the masdar, a term borrowed from Arabic linguistics, where the form this term refers to is not a verbal noun, as it is usually translated, but rather “a form which is underspecified with respect to its syntactic potential, one which can be both a complement to the predicate, i.e., it can assume nominal functions, and can be a predicate. One trait

¹⁵ Crucially, if the non-reduplicated form is polysyllabic, then the only formal difference between the two constructions illustrated by examples 52. and 53. is between active and middle voice. Polysyllabic forms can also have a habitual interpretation, however this is not their primary interpretation.

¹⁶ In fact, one speaker of the Simdega dialect even accepted the repetition of polysyllabic contentive morphemes in this construction with a habitual interpretation, a use which was however rejected by other speakers of this dialect.

which it shares with the verbal noun in Indo-European languages is that it can mark for case.” (Maas, 2007: 134, author’s translation).

Consider now the following example, where the masdar *aw-aw=te* is marked for the oblique case, as it appears in adverbial function as a secondary (i.e., “non-finite”) predicate.

54. enem	raja	ro	rani=kiyar=a?	aw-aw=te
without	king	and	queen=DU=GEN	QUAL-RDP=OBL
<i>khariya=ki</i>	<i>katij=ko</i>	<i>dher=ga</i>	<i>akbakay=ki=may.</i>	
Kharia=PL	somewhat=CNTR	very=FOC	be.in.a.flurry=MID.PST=3PL	
‘Without there being a king and queen, the Kharia panicked somewhat.’				[MT, 1:104]

As this example shows, the masdar is also not restricted to morphemes expressing an activity, as the morpheme *aw* is the purely stative qualitative predicative marker (\approx “copula”, cf. 7.3) in this example, which otherwise has the meaning ‘live; stay, remain’. In fact, masdars can be formed from any contentive morpheme, including those which can have both an activity and a stative interpretation, such as *koŋ-koŋ*, from *koŋ* ‘find out; know’, *tomon* ‘stand (up)’, *doko* ‘sit (down)’, *rusuŋ* ‘(be/ become) red’, etc.

In sum, the masdar is a form whose primary function is to express secondary (or “non-finite”) predication, although it is not restricted to this function. It is neither a noun nor a participle, as it can also be used in primary predicative function (i.e., as a matrix predicate) without any further derivational marking. The masdar is thus a “precategorical” form in its own right which in combination with the middle voice differs in meaning from the non-derived form, but which may appear in referential, attributive and predicative function.

The exceptions to bisyllabicity

Non-borrowed monosyllabic contentive morphemes belonging to the open lexical class obligatorily reduplicate to form the masdar. Borrowed morphemes, however, do not, except for *col* ‘go’, which is unique in other ways as well, being, e.g., together with *dɛl* ‘go’ one of the two morphemes in the language in which a morpheme-final /l/ is deleted before a morpheme-initial /n/ in an enclitic (cf. Section 2.4). The fact that it is the only borrowed morpheme in the language which reduplicates strongly suggests that this is due to influence from *dɛl* ‘come’ here as well.

Now consider now the following six forms, some of the very few monosyllabic contentive morphemes which do not reduplicate in a Case-syntagma:

55. *laŋ* ‘tongue; language’ *moʔ* ‘smoke’ *moʔd* ‘eye’ *seʔ* ‘louse’
 tiʔ ‘hand’ *oʔ* ‘house’

At first glance, it might seem that these morphemes are not reduplicated for the simple reason that they are already “nouns”, whereas monosyllabic “verbs” need to be reduplicated to function as “nouns”. However, as will now be shown, examples such as those in (55) are rare exceptions—and generally also highly common forms.

In order to determine the extent to which such monosyllabic forms occur in Kharia, a survey was conducted of the contentive morphemes from a Kharia-English lexicon (Peterson, 2006, vol. 3), comprising all non-derived contentive morphemes belonging to the open class under the entries *a*.*e* which do not appear to have been borrowed from Indo-Aryan or other sources or, if borrowed, show signs of having been borrowed at a very early date¹⁷ (comprising altogether 84.5 pages or considerably more than 25% of the entire lexicon at that time). The results of this survey are shown in Table 4.1.

Table 4.1: Statistics for inherently polysyllabic, reduplicating and inherently monosyllabic contentive morphemes of the open class

	<i>Total number</i>	<i>% of Total</i>
Polysyllabic contentive morphemes (always):	229	83%
Monosyllabic contentive morphemes which reduplicate in a Case-syntagma:	35	12.7%
Monosyllabic contentive morphemes which do not reduplicate in a Case-syntagma:	12	4.3%
	<i>276</i>	<i>100%</i>

¹⁷ In fact, if loan items had also been considered, this would probably have decreased the percentage of the inherently monosyllabic contentive morphemes even further, since there is an abundance of bi- and trisyllabic loan morphemes which do not reduplicate in a Case-syntagma. These were not taken into consideration in this survey. Causative-marked semantic bases were also not taken into account, which never reduplicate, as they are always at least bisyllabic. Had these been included, the percentage of exceptions would have been exceedingly small.

Note that the class of inherently monosyllabic contentive morphemes is small (under 5%) and can therefore be viewed as an exception to the general rule. It is true that this class includes many forms which denote physical objects, as in the following, where “*” denotes a possible loan word:¹⁸

56. *ba?* ‘rice paddy’ **ban* ‘spell, charm’ **chāi* ‘female pig’
 bo? ‘place’ (n.) *bi?d* ‘basket’ *ḍa?* ‘water; culture; waist’

First note that there is also one item contained in this figure of 4.3% which does not refer to a physical object and which, incidentally, may also translate as an adjective: *bay* ‘craziness; crazy’ Others, such as *ban* ‘spell, charm’, also do not refer to a physical entity. One further item is merely an alternative form of a bisyllabic contentive morpheme with the same meaning: *ber* / *behar* ‘who?’ Another form included in this figure is actually no longer used as an independent morpheme by speakers of these two dialects but is only found in “compounds” (cf. Section 4.6).¹⁹

57. *ḍay* ‘woman’, cf. *ol-ḍay* [take-woman] ‘marry’,
 kulam-ḍay [sibling-woman] ‘sister’

Taking these facts into account, the number of monosyllabic free-forms which do not reduplicate in a Case-syntagma and which denote a physical entity now drops to 8 out of 276 forms, or approximately 2.9%, a very small minority of forms which can safely be considered exceptions to the general rule, i.e., the large majority of contentive morphemes, whether referring to physical objects, humans, animals or events/situations, are either polysyllabic or reduplicate in the case-syntagma.

We occasionally also find reduplication with “typical nouns”: Cf. the forms *ḍuḍuḍ* ‘eel’ and *saḥsaḥ* ‘turmeric’ Although neither of these appears to still be used in its monosyllabic form, *saḥ* is given in Floor et al. (1934: 111)²⁰ in its monosyllabic form and it would seem that until recently it was used with the meaning ‘become yellow’ and as a reduplicated

¹⁸ As I have not been able to locate a Sadri-English dictionary (only English-Sadri and Hindi-English), I cannot rule out that some of these forms are in fact loan words. “*” is only used here where there is some reason for doubt.

¹⁹ In the more southerly dialects, spoken in and near Orissa, *ḍay* appears to be used as a free-form but not in other dialects.

²⁰ This form is given there as *sang* and is said to derive from *sarsang*, *sarson* (?) “= which has a yellow flower. yellow.” I am not familiar with either of these forms, neither of which is listed separately in Floor et al. (1934).

form with the meaning '1. yellow; 2. turmeric' 'yellow' is now expressed by the Sadri loan word *piyar*. Similarly, *ḍuṅḍuṅ* 'eel' no longer appears to be used productively in its non-reduplicated form, **ḍuṅ*, although this form is found in old compounds, such as *kaḍoṅ* 'fish' (Pinnow, 1959: 77, §64).

There is furthermore no reason to assume that these exceptions are inherently nouns, as at least those which have been tested from this group can also be used predicatively. For example, *moʔ* 'smoke' used predicatively means '(of smoke) to emerge' in the middle voice (*moʔ=ki* [smoke=MID.PST] 'smoke came out [of X]') while the active has the meaning 'cause smoke to emerge (from something)' (*mokh=oʔ* [smoke=ACT.PST] 's/he caused smoke to come out [of X]').²¹

There is also a possible diachronic explanation for some of these forms: Consider for example the monosyllabic forms *beʔḍ* 'son' and *biʔj* 'daughter' (neither included in this figure, as they derive from Indo-Aryan), which are only found when followed by a marker of inalienable possession, e.g. *beʔḍ=nom* [son=2poss] 'your son', *biʔj=nom* [daughter=2poss] 'your daughter'. Otherwise, these contentive morphemes have the forms *beḥa* and *beḥi*, respectively. This may provide a diachronic explanation for why three of the forms given in the list above, *laṅ* 'tongue', *moʔḍ* 'eye' and *tiʔ* 'hand', are monosyllabic, assuming that body parts were once generally (perhaps obligatorily) marked for inalienable possession, which would appear to be declining in use (cf. e.g., Abbi, 1993: 551), although this is uncertain.

Summary—the masdar

In sum, the data indicate that reduplication in the masdar, while derivational, does not derive "nouns" or "participles" from "verbs". Rather, although its primary function is to derive a bisyllabic phonological word from a monosyllabic form which can then be used in attributive or referential function, it is not restricted to this use and is also found in primary predication. It is thus best considered a process which derives "pre-categorical" bisyllabic forms from "pre-categorical" monosyllabic forms so that these can stand alone as phonological words. Also, while there are a number of monosyllabic native morphemes which do not reduplicate in these environments, these make up only ca. 4.3% of the native lexicon and are exceptions to the general rule, most likely due to a combination of diachronic developments and their relatively high frequency. Also, not

²¹ 'to smoke', with the English meaning of 'smoke [e.g., a cigarette]', is expressed with *uʔḍ* 'drink', the general means of expressing this in South Asia. I.e., in most South Asian languages, one does not 'smoke' a cigarette but rather 'drinks' it.

all of them refer to physical entities, and all would appear to be compatible with any discourse function.

4.3.3 *The -NV- infix*

The infix *-NV-* derives contentive morphemes from other contentive morphemes; the meaning of these derived forms differs from that of the underived forms, although unlike the reduplication discussed above, the semantic connotations here are unpredictable and idiosyncratic.

In this infix *-V-* has the same quality as the vowel preceding *-N-* in all cases other than the slightly irregular *mo<ne>n* ‘one each’ from *moŋ* ‘one’ Otherwise, the only irregularity in the formation of this category is in the pair *mui* ‘dream (v.)’ / *mu<nu>* ‘dream (n.)’ *-N-* is a nasal of indeterminate quality. It is usually realized as /n/, but /m/ is also occasionally found.²² Table 4.2 presents a few examples. A more exhaustive list is given as the first entry under **n** in Peterson (2009).

Table 4.2: The “nominalizing” infix *-NV-*

<i>bel</i> ‘spread out (a mat or bedspread)’	<i>be<ne>l</i> ‘bedding’
<i>biɾay</i> ‘test’ (v.)	<i>bi<ni>ɾay</i> ‘test’ (n.)
<i>bui</i> ‘keep, raise (an animal)’	<i>bu<nu>i</i> ‘pig’
<i>ɬeʔj</i> ‘chop’	<i>ɬe<ne>j</i> ‘(medium-sized or large) hatchet’
<i>gil</i> ‘beat’	<i>gi<ni>l</i> ‘beating’
<i>jib</i> ‘touch’ (v.)	<i>ji<ni>b</i> ‘touch’ (n.)
<i>joʔ</i> ‘sweep’	<i>jo<no>ʔ</i> ‘broom’
<i>kol</i> ‘count’	<i>ko<no>ol</i> in <i>lebu ko<no>l</i> ‘census’ (cf. <i>lebu</i> ‘man, person’)
<i>kuj</i> ‘dance’ (v.)	<i>ku<nu>j</i> ‘dance (event)’ (n.)
<i>lebui</i> ‘love’	<i>le<ne>bui</i> ‘love’ (HɾA:92)
<i>moŋ</i> ‘one’	<i>mo<ne>n</i> ‘one each’ (HɾA:146)
<i>mui</i> ‘dream’	<i>mu<nu></i> ‘dream’
<i>muʔsiŋ</i> ‘rise (of the sun)’	<i>mu<nu>ʔsiŋ</i> ‘east (i.e., where the sun rises)’
<i>rab</i> ‘bury’	<i>ra<na>b</i> ‘burial ground’ (HɾA:161, notes)
<i>roʔ</i> ‘drop, spill’	<i>ro<no>ʔ</i> ‘dropping or spilling’ (HɾA:160)
<i>tol</i> ‘tie’	<i>to<no>l</i> ‘string, shoe-lace’
<i>uʔphe</i> ‘three’	<i>u<nu>ʔphe</i> ‘three each’ (HɾA:146)

²² This infix can be traced back to Proto-Austro-Asiatic, as it is also found in languages of the Mon-Khmer group, e.g., (marginally) in the Southern Aslian (Mon-Khmer) language Semelai, spoken in Malaysia (Kruspe, 2004: 86). For a discussion of this marker from a broader Munda perspective, see Zide (1968).

In some cases the simple, non-derived form no longer exists but can be assumed for some earlier stage:

58. * <i>ga²d</i> ‘reap’	<i>ga<na>²d</i> ‘sickle’
* <i>kon</i> ‘small’	<i>ko<no>n</i> ‘(become) small’
(*) <i>sel</i> ‘pray’	<i>se<ne>l</i> ‘prayer’
(*) <i>si</i> ‘plow’ (v.)	<i>si<ni></i> ‘plow’ (n.)

There is no contentive morpheme **ga²d* in Kharia, but such a contentive morpheme is found in the South Munda language Sora with the meaning ‘cut’. Also, although **kon* is not a contentive morpheme in the modern language, there are numerous cases in which it is found in erstwhile compounds: *konsel* ‘young woman’, *kondu?* ‘child’.

The case with *sel* and *si* is somewhat different: *sel*, with the meaning ‘pray’, is still found in the texts in Pinnow (1965a), while speakers who we questioned all rejected the form and use *senel* both predicatively (‘pray’) and referentially (‘prayer’). Similarly, *si* is given in Biligiri (1965: 194) with the meaning ‘plow’ (v.), whereas nowadays this is realized as *silo?*, where *-lo?* is a compound form (cf. *uslo?* ‘land’, *lo?kha* ‘earth, soil’; see Section 4.6.2).

With the exception of numerals, from which this derivational strategy apparently once derived distributives (cf. *mo<ne>n* ‘one each’ and *u<nu>?phe* ‘three each’ in Table 4.2), the derived forms all translate as nouns in English which are related to the underived contentive morpheme in a rather idiosyncratic way: The derived forms of *bel* ‘spread out’ and *bui* ‘raise (an animal)’ denote patients of the respective action, the thing spread out (*benel* ‘bedding’) and the animal raised (*bumui* ‘pig’). On the other hand, the derived form is the instrument of the event denoted by other contentive morphemes, such as *de²j* ‘chop’ or *jo?* ‘sweep’ (*dene²j* ‘hatchet’ and *jono?* ‘broom’) or an abstract concept, as with *gil* ‘beat’, *jib* ‘touch’ and *ro?* ‘spill’ (*ginil* ‘the act of beating’, *jimib* ‘touch (n.)’, *rono?* ‘the act of spilling’).

This infix was undoubtedly originally used to derive bisyllabic forms from monosyllabic morphemes, similar to the reduplication discussed in the previous section, so that these forms could stand alone as a phonological word. Whereas in the modern language monosyllabic forms are consistently reduplicated in these environments, the older language seems to have preferred the *-NV-* infix. However, forms marked by this infix in the modern language are best considered contentive morphemes in their own right.

With respect to bisyllabicity: Note that only three of the examples given in the list above, *bi<ni>ɽay* 'test', *le<ne>bui* 'love' and *u<nu>ʔphe* 'three each', derive from bisyllabic morphemes.²³ The first of these examples is somewhat dubious, as in the same story from which this example derives, the form *biɽay* is also found used as the base of a genitive attribute: *biɽay=aʔ samay* [test=GEN time] 'time of testing' (Pinnow, 1965a: 57). The second form is at least equally dubious, as Pinnow (1965a: 93, note 47) notes that the speakers he asked to read this text aloud all read the form <*lenebui*> in the manuscript simply as *lebui* 'love', which is its usual form. Speakers we consulted rejected both *bi<ni>ɽay* and *le<ne>bui*. These two forms are therefore considered here to be either dialectal or perhaps idiolectal variants. Only the form *u<nu>ʔphe* 'three each' remains to be accounted for, and it would seem that we are dealing here with a secondary function of this infix, that of deriving distributives from cardinal numbers, a construction which is no longer productive (both these forms are from older texts (Pinnow, 1965a: 146)). We therefore conclude that -*NV*- was used to derive bisyllabic phonological words from monosyllabic contentive morphemes.

With respect to parts of speech: As noted above, in the modern language this infix simply derives different contentive morphemes from an underlying contentive morpheme and we cannot assign any one function to it. This process is also opaque to speakers who we consulted, who saw no connection between the underived and derived forms and even rejected our interpretations of them.

Like any other contentive morpheme, we should expect that these morphemes can also be used predicatively. Although it has not yet been possible to check this interpretation with native speakers for all of these morphemes, those which could be checked were indeed all perfectly compatible with a predicative function, cf. *ginil* 'beat (HABITUALLY)', *tonol* 'tie', etc., and *ganaʔd*, *benel* and *deneʔj* may all also be used attributively. Although it was not possible to go into the semantics of these morphemes in predicative function for reasons of time, for a few this was possible, and the picture which emerges is not entirely clear: E.g., while *ginil* was

²³ Although *mu<nu>ʔsiŋ* 'east' in Table 4.2 derives from the bisyllabic *muʔsiŋ*, *muʔsiŋ* is itself originally a compound, consisting of *muʔ* 'emerge' and the compound form of *siŋkom* 'star', which seems to have originally meant 'sun' (cf. e.g. *siŋi* 'sun' in the South Munda language Remo (Bhattachary, 1968: 127; Fernandez, 1967: 154) or *siŋi* 'the sun; daylight' in the North Munda language Ho (Deeney, 1978: 319)). Thus, from a historical perspective at least, we have here a compound formed by **mu<nu>ʔ* 'emergence' (?), from the monosyllabic contentive morpheme *muʔ* 'emerge', and **-siŋ* 'sun'

interpreted as a habitual, together with the middle voice, which would make it a kind of alternative to the masdar (despite the fact that speakers generally claim not to be aware of any relation between the infix and non-infix forms), the speaker consulted on *tonol* was unsure as to how *tonol* differs from *tol*, suggesting that *tonol* might mean ‘re-tie’ (< HABITUAL?). Similarly, *benel* was given with the same meaning as *bel*, although upon being asked, the speaker agreed that it could also mean ‘become a bed-spread’ Further work is necessary here, nevertheless even these preliminary data make it clear that the infix form, whatever its exact semantics, is not an “inherent noun” but rather a kind of (derived) contentive morpheme which, like all other contentive morphemes, may be used in all functions.

In summary, although these are clearly derived forms, and although this category is neither productive nor predictable in its semantic relation to the underived form, this infix is not a nominalizer: It derives one lexical concept from another in a highly unpredictable way, but not “nouns” from “verbs” The only predictable component here is that the resultant form is in some way connected to that of the underived form, such as *joʔ* ‘sweep’ and *jonoʔ* ‘broom’ Otherwise, the two are merely different contentive morphemes.

4.3.4 -NV- + Reduplication

This construction, discussed in Pinnow (1965a: 161, note), combines the infix *-NV-* just discussed with the reduplication of the (underived) contentive morpheme, typical of masdars. Instead of the infix *-NV-*, *-NVm-* is also found, in which case the consonant in the coda of the underlying contentive morpheme is deleted. In one case, it has the form *-Vm-* (cf. *delem-dol*). Consider the following forms. The first three, with the exception of the (dialectal) form *delem-dol*, are from Pinnow, the remainder from this author’s own fieldwork:

59. Underived morpheme	Derived form	
<i>del</i> ‘come’	<i>de<nem>-del</i> , <i>del-dol</i>	‘arriving
<i>goʔj</i> ‘die’	<i>go<nom>-goʔj</i>	‘deceased (person)’
<i>raʔb</i> ‘bury’	<i>ra<na>ʔb-raʔb</i>	‘burial ground; grave’
<i>col</i> ‘go’	<i>co<nom>-col</i>	‘about to go’
<i>kol</i> ‘count’	<i>ko<no>l-kol</i>	‘population’

It is extremely difficult to determine the status of this form. It does not appear to have ever been very productive and the five forms given here are the only attested forms to date, as compared with at least 45 occurrences of the *-NV-* infix alone. Also, there seems to be no clear function assignable to these forms. For example, *qe<nem>-qel*, in its dialectal form *qelem-qol*, is a variant of the regular imperfective “converbal” form *qel=ga qel=ga* ‘arriving’ in the dialect of one southern speaker we worked with:

60. *la?* *aniŋ=a?* *khariya* *bulbul*, *yane* *babilon*, *po?da* *tay*
 then 1PL.INCL=GEN Kharia Babylon i.e. Babylon village ABL
- mu?=kon* *qel-qol* *arlo?*, *serlo?siŋ* *ghay* *hoy=kon*
 emerge=SEQ come-? north south way become=SEQ
- u=ti?j* *qel=ki=may*.
 this=side come=MID.PST=3PL
- ‘Then our Kharia [ancestors], having left Bulbul, i.e., Babylon, *coming along*, via the north and south, came to this side [i.e., here].’ [MT, 1:20]

qe<nem>-qel was also cited by this same speaker in the following use:

61. (*ho=kaŋ*) *qe<nem>-qel* *ayi?j*. ‘S/he will come.’
 that=SG.HUM come-? QUAL.PRS

This is not a productive means of referring to future events and seems to be a feature of the southern dialect of this speaker. Further research is necessary. The same speaker also cited the form *co<nom>-col*, with a similar function, which however was not accepted by other speakers consulted.²⁴

It is worth noting in this respect that, with the exception of *ko<no>l-kol* and the forms *qe<ne>m-qel* and *co<no>m-col* for this one speaker, none of the forms given in the list above are in common use and none of the younger speakers consulted were familiar with any of them.

²⁴ I worked with this speaker on my first visit to Jharkhand in 2001. Unfortunately, I could not locate him during my subsequent four visits to investigate this issue further.

4.3.5 *Derivation and parts of speech—A summary*

In sum, we have seen that, with the exception of the small, closed class of proforms and deictics of time and place, virtually all underived contentive morphemes in Kharia may be considered “precatatorial” in the sense that they may appear in referential, attributive and predicative function, and the present section has shown that there is no reason to assume that this is any different for derived forms, i.e., there do not appear to be any “nominalizers”, “adjectivizers” or “verbalizers” in Kharia.

Although precatatoriality in this sense is perhaps a necessary condition for a language which does not have a noun/verb distinction, it is for many researchers still not a sufficient criterion: Even if it could be shown that ALL contentive morphemes in Kharia are precatatorial, it could still be argued that this merely means that any contentive morpheme can appear as a noun or a verb (etc.), not that these classes do not exist. As we shall see in the following section, Kharia provides still further evidence against assuming the presence of nouns and verbs as lexical classes: In addition to the precatatoriality of contentive morphemes, the units on which predicates, referential units and attributes are based are in fact not lexical categories at all but rather syntactic structures. In other words, lexical classes in Kharia play at best a marginal role in the morphosyntax.

4.4 “NPs” AS VERBS? SYNTACTIC CATEGORIES, NOT LEXICAL CATEGORIES

To begin with a simple example, note that there are many predicates in Kharia (and Munda languages in general) which appear to be typical “verbs” but whose “lexical head” consists not of one but of two contentive morphemes:

62. <i>musniŋ</i>	<i>jhaŋi</i>	<i>kõpuruʔ=ki</i>	<i>jume</i>	<i>katiʔb=kon</i>	<i>socay=oʔ=ki</i>
one.day	all	man=PL	assemble	gather=SEQ	think=ACT.PST=PL
<i>no</i>	<i>“ho</i>	<i>dano=te</i>	<i>i=gud=ga</i>	<i>tar</i>	
CMPL	that	demon=OBL	what=like=FOC	kill	

oʔ-gur=e=niŋʔ”

CAUS-fall=ACT.IRR=1PL.INCL

‘One day, all the men gathered (= *assembled [and] gathered*) and thought “How will we kill (= *kill [and] cause to fall*) that demon?”’

[MT, 1:66]

Example (62) is typical of a large class of complex finite predicates. These are usually considered to be verbs whose lexical head consists of a compound. However, as will be shown in Sections 6.3.2 and 6.6.1, these are in fact not compounds as the two (or more) contentive morphemes can be separated by conjunctions such as *ro* 'and' and as the non-final semantic bases can also be partially finite.

If these complex heads are not compounds but simply juxtaposed contentive morphemes, the question then arises as to whether other types of complex semantic heads are possible, which is clearly the case. To begin with, consider the following example, where what would appear to be "nouns" and "pronouns" marked for the genitive case can also serve as the "lexical" head of a predicate, with no derivational morphology. This is an entirely productive process.

63. a. *ayo=ya?* *keciya* b. *ayo=ya?=yo?*
 mother=GEN money mother=GEN=ACT.PST.1SG
 'mother's money' 'I made [it] mother's.'
- c. *ayo=ya?=ki.*
 mother=GEN=MID.PST
 'It became mother's.'
64. a. *ijn ho=ka?te* *ijn=a?=yo?*
 1SG that=SG.HUM=OBL 1SG=GEN=ACT.PST.1SG
 'I *adopted* him/her (i.e., I *made* him/her *mine*).'
- b. *am ho=ka?te* *am=a?=yo?*
 2SG 3=SG.HUM=OBL 2SG=GEN=ACT.PST.2SG
 'You *adopted* him/her.'

Examples (63) and (64) strongly suggest that any analysis of these predicates as "verbs", i.e., as a lexical category, must be viewed with considerable scepticism, as we would then have a "verb" consisting of a case-marked "noun". As we shall presently see, however, this is by no means the only problem with such an analysis.

In fact, what appear to be entire "NPs" can serve as the "lexical" head of a predicate in Kharia. As the term "lexical" is clearly not appropriate for such structures, this unit will henceforth be referred to exclusively by the more neutral terms "semantic head" or "semantic base". For example, a "noun" with a quantifier, as in (65), or a demonstrative, as in (66), can also serve as the semantic head of a predicate. If we claim these are (zero

derived?) verbs, it is not immediately obvious how to describe the structure of these verbs.

65. *ubar rocho⁷b=ki=jɹ*
two side=MID.PST=1SG

‘I moved to both sides (i.e., this way and then that).’

66. *ho rocho⁷b=ki=jɹ*
that side=MID.PST=1SG

‘I moved to that side.’

“NPs” containing a genitive attribute can also be used predicatively, as in (67) and (68), which would also seem to indicate that the head of the “verb” is actually the head of a noun phrase. Working with “nouns” and “verbs” thus creates a number of problems which cannot easily be accounted for, at least not without “hidden verbs” or other non-overt categories.²⁵

67. *hoɖom dimu=jɔ kongher=te dho?⁷na=ya? kornis=o?*
other day=ADD boy=OBL grab=INF=GEN attempt=ACT.PST

muda um pa⁷=o?
but NEG be.able=ACT.PST

‘The next day as well [the witch] *tried to grab* the boy but could not.’ [BB, 1:52]

68. *bharat=ya? lebu=ki bides=a? lebu=ki=ya?*
India=GEN person=PL abroad=GEN person=PL=GEN

rupraŋ=ki=may.
appearance=MID.PST=3PL

‘The Indians *took on the appearance of foreigners* (e.g. by living abroad so long).’

These examples show that we cannot simply speak of nouns and verbs in (Simdega) Kharia in the usual sense because TAM/PERSON-syntagmas and Case-syntagmas in Kharia are not formed in the lexicon but rather in the

²⁵ However, even if we do assume here a “hidden verb”, the analysis remains problematic in some cases. See below, beginning with example 75..

syntax. This analysis fits in well with the enclitic nature of TAM, person/number/honorific status, and case marking (cf. Chapter 3). Thus, both the components of the semantic base of the TAM/PERSON-syntagma or the Case-syntagma as well as these functional markers are syntactic atoms, not affixes, and the enclitic functional markers merely attach to the last element of this syntactic “phrase”, regardless of its status.

Considering again the structure of examples (63)–(68), we see that the structures in (63) and (64), in which a genitive-marked “noun” serves as the semantic base of a TAM/PERSON-syntagma, are in fact also syntactic: Here, the lexical heads (e.g., *keciya* ‘money’ in (63) and *beṭa* ‘son’ or *beṭi* ‘daughter’ in (64)) have simply been omitted as their identity is clear from the context, a very common strategy in Kharia. In other words, if the lexical head of a syntactic construction is recoverable from context, it may be omitted. We will return to this topic in Chapter 5.

With some “nominal” predicates with a genitive-case modifier of the type described above it is possible to analyze these as omitting a postposition which takes (or can take) the genitive, as in (69). This could then be considered a kind of obligatory but oblique “object” which is usually marked by a postposition:

69. *ho=kaṭ* *iŋ=a?* (*thon / ghaʔd*) *madet=yo?* ‘S/he helped me.’
 that=SG.HUM 1SG=GEN for for help=ACT.PST

However, this is not always possible, and the inclusion of a postposition governing the genitive is in fact often ungrammatical, as (70) shows:

70. *iŋ* *ho=kaṭ=a?* (**thon / *ghaʔd*) *main=oʔj* . ‘I honored him/her.’
 1SG that=SG.HUM=GEN for for honor=ACT.PST.1SG

As the fact that a postposition cannot always be used in these expressions, as in (70), clearly rules out deriving these genitive-marked phrases from “underlying” postpositional phrases in general, the only viable alternative is to consider the genitive-case marked element to be a modifier of the final element of the semantic base: If two elements are in a possessive relationship to one another in Kharia, then the possessor may appear as a genitive-marked modifier to the possessee, regardless of whether the resultant syntactic construction appears in predicative or referential function. This structure RESEMBLES an NP in English and most other languages, however it would make little sense to consider it an NP in Kharia, as this

would require that the head of the (endocentric) NP construction be a noun and that the entire NP would then have to be (zero) converted—into the syntax!—into a verb in order to function as a predicate.

Not only do we find TAM/PERSON-syntagmas whose semantic heads are marked for referentiality, as in (66), they may also be marked for inalienable possession and number: Although these markers are typically found with Case-syntagmas, the following examples show that they may also form part of the semantic head of a TAM/PERSON-syntagma.²⁶

71. *boksel=nom* *go²d=ki*. ‘She became your sister-in-law.’
sister.in.law=2POSS C:TEL=MID.PST

72. *ho=je?* *u=je?=ki* *go²d=ki*.
that=SG.NHUM this=SG.NHUM=PL C:TEL=PL

‘That became these (= that “these-d”).’

“Case” marking other than the genitive, however, is not compatible with TAM/PERSON marking. Thus, with the exception of the genitive, which does not play a role at the clause level, case marking (including adpositions) is the functional head of a Case-syntagma.

73. **sahar=te=ki=jL*
city=OBL=MID.PST=1SG

Furthermore, just as both TAM/PERSON-syntagmas and Case-syntagmas are formed in the syntax in Kharia, “modifiers” are also formed in the syntax, being in fact either TAM/PERSON-syntagmas or Case-syntagmas. The following presents one simple example. For further, more complex examples, see Section 7.6 (“relative clauses”).

74. *col=ki* *memon*
go=MID.PST year
‘last year (= *it=went* year)’

²⁶ *go²d* is a v2 and simply denotes telicity (6.5.3). It is not a verbalizer and is also compatible e.g. with typical action-denoting contentive morphemes, cf. *col go²d=ki* [go C:TEL=MID.PST] ‘s/he left’.

Finally, examples such as the following show that assuming a “zero” or “ellipsed verb” is not only uneconomical and forces a pre-conceived notion onto Kharia, it is also questionable as to whether such an analysis is really feasible. Consider the following Case-syntagma:

75. *iku²d jughay du?kxo buŋ*
 very much sorrow INST
 ‘with great sorrow’

Note that it is possible to insert some element with the meaning ‘become’ after the instrumental postposition *buŋ*:

76. *iku²d jughay du?kxo buŋ hoy go²d=ki*
 very much sorrow INST become C:TEL=MID.PST
 ‘S/he became very sad (= with much sorrow).’

However, this cannot serve as a general explanation for such complex predicates in Kharia, as it is not possible to “ellipse” *hoy* in (76):

77. **iku²d jughay du?kxo buŋ go²d=ki*

The reason this is so is that, as stated in Section 4.1, markers of a Case-syntagma (i.e., case and adpositions) may not combine with markers of the TAM/PERSON-syntagma, as a constituent may not simultaneously be a TAM/PERSON-syntagma and a Case-syntagma. Instead, we find structures such as the following, with a “mandatory reordering” of the respective elements:

78. *gupa lebu du?kxo buŋ iku²d jughay go²d=ki*
 guard person sorrow INST very much C:TEL=MID.PST
 ‘The shepherd *became very depressed*.’ [RD, 1:18]

However the structure in (78) is best analyzed from a theoretical perspective, it clearly shows that assuming the presence of “hidden verbs” to account for these structures in Kharia is at best problematic, unless this non-overt unit is allowed to have a unique distribution among all morphemes in the language.

In view of these facts, it is most “economical” in the Simdega dialect of Kharia to view TAM/PERSON-syntagmas and Case-syntagmas as consisting of a semantic head, potentially containing deictic, anaphoric and lexical

information, and a functional head, containing either case marking (= Case-syntagma) or TAM and PERS/NUM/HON-marking (=TAM/PERSON-syntagma). IT IS THE STATUS OF THE FUNCTIONAL HEAD ALONE WHICH DETERMINES THE INTERPRETATION OF A CONSTITUENT IN A CLAUSE AS EITHER A TAM/PERSON-SYNTAGMA OR A CASE-SYNTAGMA, AND BOTH TYPES OF SYNTACTIC CONSTRUCTIONS CAN TAKE THE SAME TYPES OF SEMANTIC HEAD.

The only requirement is that a TAM/PERSON-syntagma and a Case-syntagma have 1. a functional head and 2. some kind of semantic head. The semantic head may be a simple contentive morpheme, a genitive-marked element, a demonstrative, a quantifier, or any combination of these elements, and may be marked for inalienable possession and number.

On the other hand, for the Gumla dialect an analysis in terms of nouns and verbs in the traditional sense seems preferable, although a large portion of the lexicon is nevertheless underspecified for lexical class (see next section).

4.4.1 *The question of acceptability*

At this point, the reader may justifiably be wondering whether the examples given above, in what appears to be an NP functioning as a verb, are at all “natural” in spoken or written Kharia or even whether they truly are grammatical, given the fact that most of the examples cited are from interviews. And in fact, not all speakers accepted all examples.

In general, the same comments apply here as with respect to the acceptance by speakers of contentive morphemes denoting physical and animate entities in predicative function (4.2) or the use of polysyllabic contentive morphemes in the marked function of the middle voice (4.3.2): Younger speakers of the Simdega dialect generally accepted such complex TAM/PERSON-syntagmas—here the rate of acceptance was virtually 100%—while speakers of the same age of the Gumla dialect tended to reject these. Furthermore, younger speakers were generally more accepting of such structures than older speakers, and men in general somewhat more than women.

With respect to the age difference and the degree to which examples of these types were considered grammatical or not, this appears to be due to non-linguistic factors: All elder speakers consulted, with one exception, taught Kharia grammar and literature and were immensely interested in using only “proper” or “pure” (*sudh*) Kharia.²⁷ It would seem that this “pure” version of the language is not that which is actually spoken but

²⁷ The one elder speaker who did not teach Kharia had nevertheless taught most of his life at a local (English-medium) school. Thus, many of the following comments would appear to apply to him as well.

rather a literary variety which is highly influenced by Hindi and Sadri, languages which have clearly definable lexical classes,²⁸ and constructions of the type given in the previous section are not possible in these languages. We assume that it is precisely this Indo-Aryan influence, which is evident in virtually all aspects of Kharia grammar (cf. Abbi, 1993; 1997; Malhotra, 1982, Chapter 7), which led these elder speakers to reject such examples.

This hypothesis is also supported by comments many of the younger speakers themselves made. For example, younger speakers often stated that although a certain construction (such as the ones in question here) are not “pure Kharia”, they are used, especially by younger speakers, although their use should be avoided.²⁹

This attitude towards “pure” Kharia may explain why women under 40 were slightly less likely to accept such examples as grammatical, as almost all women in this age group who this author worked with were planning on becoming teachers after graduation, generally Kharia teachers in elementary schools. In contrast, only one of the men who this author consulted was planning on such a career. This is of course highly speculative and the topic awaits further study but at present it seems to be the most plausible explanation for the distribution of the data.

Further research on the factors determining the acceptability of these constructions for individual speakers is necessary. But whatever the reasons for these differing views may be, the following attested examples from the literature would seem to imply that negative replies in interviews may partly be due to the interview situation itself, as they show that these constructions do in fact occur in “natural” speech and writing. In addition, they also show that these constructions cannot be new to the language nor are they restricted to younger speakers, as most of these are either from Pinnow’s texts (1965a) or from plays written by the highly respected, now deceased Khrist Pyari Kerketta (Kerkeṭṭā, 1990):

*Genitive attribute or other attribute in the semantic head of the TAM/
PERSON-syntagma:*

79. *ij̄n lere?=ta=ij̄n no ponomosor tunbog=a? ghad̄*
1SG joy=MID.PRS=1SG Cmpl God daytime=GEN for

²⁸ Cf. e.g. Montaut (2004) for a discussion of lexical classes in Hindi.

²⁹ Interestingly, similar comments were also made on the use of the Past II in =*kho?* which derives from the past perfect (6.4.1.1). This construction is especially common with younger speakers but is highly frowned upon by most elder, educated speakers.

patar=a? upay=o?.

light=GEN means=ACT.PST

'I am happy that God *managed to make light* for the day.' [HJPa:96]

80. *i ponomosor jhaṛi cij=a? bolan=o? ?*

Q God all thing=GEN thought=ACT.PST

'Did God *think of everything?*' [HJPa:96]

81. *ponomosor baʔj=ta no moṇ ubar katib katib*

God like=MID.PRS CMPL one two assemble REP

aw=ta=ki ho=ki hodom=ki=ya? thom jahã

QUAL=MID.PRS=PL that=PL other=PL=GEN for INDEF

lekhe=ya? sanḡhar=e=ki.

kind=GEN help=ACT.IRR=PL

'God likes it when one or two people who are together help others somehow (= *some kind of help*).' [HJPa:103]

82. *jati=ya? sewa[=e]=iṇ, cahe ba? rumkub*

ethnic.group=GEN serve=ACT.IRR=1SG whether paddy rice

co=na, main co=na,

go=MID.IRR honour go=MID.IRR

'I will *serve the Kharia* (= [*our*] *ethnic group*), whether I run out of food, lose my honour, [or] [Kerkeṭṭā, 1990: 10]

83. *ele goṛho=wa? boiri=ki=te aḡ=ki=ya?*

1PL.EXCL fort=GEN enemy=PL=OBL ANAPH=PL=GEN

goṛho tay dḡheir dḡisa? dḡoḡ=na=? kornis=e=le.

fort ABL much far take=INF=GEN attempt=ACT.IRR=1PL.EXCL

'*We will attempt to take* the enemy of the fort very far away from their fort.'³⁰ [Kerkeṭṭā, 1990: 17]

The following example, while not from a narrative or written text, was cited as an example for the use of the morpheme *lo?* 's:ITER'. The speaker

³⁰ For a further, similar, non-elicited example, cf. 67. above.

here wanted to give an example for a situation where *kosor* ‘dry (up)’ and *lo?* ‘S:ITER’ could appear in combination.

84. *ijn* *hoḍom* *kamu=na* *la?=khoʔj.* *jur-jur* *ba?*
 1SG other work=INF IPFV=PSTII.1SG spread.out-RDP rice

kosor lo?=ki.

dry S:ITER=MID.PST

‘I was *doing some other work*. The rice which had been spread out [to dry] was slowly drying.’

TAM/PERSON-syntagmas with complex semantic heads which generally appear in referential function:

85. *botoj=ta=pe* *ho=ki* *lutui* *su=kon* *pe?* *cakhna?*
 fear=MID.PRS=2PL that=PL clothes put.on=SEQ rice curry

iʔjthan *kinbhar=na=pe,* *ijn=ko* *laʔe=na=ijn.*

cowdung courtyard=MID.IRR=2PL 1SG=CNTR fight=MID.IRR=1SG

‘Those of you who are afraid, you put on your [house] clothes and see to your house work like cooking and cleaning (=rice [and] curry [and cleaning the] courtyard [with] cowdung), but I will fight.’

[Kerkeṭṭā, 1990: 7]

86. *yo=na* *yona* *eʔheʔd* *goḍ=ki* *bhere* *modi=ya?*
 see=INF REP annoy C:TEL=MID.PST time Modi=GEN

ma=ḍom *timson* *soṅol* *goṅṭiṅ=na* *col=ki.*

mother=3POSS fire firewood cook.rice=INF go=MID.PST

‘When she got annoyed at watching [them fight], Modi’s mother went to gather firewood for cooking (= she went to *fire firewood cook.rice*).’

[RD, 2:55]

TAM/PERSON-syntagmas with a complex semantic head containing two “nouns”, both of which are modified by a demonstrative:

87. *māḡta* *tamaku* *ragday=taʔj,* *bādhna* *kaʔiʔj* *saʔrakhi* *lekhe*
 Mangta tobacco rub=MID.PROG Bandhna little stubborn like

han=tiʔj u=tiʔj khor=ta, kaʔiʔj jorsāy
 that=side this=side ITER³¹=MID.PRS little strongly

bul=siʔ.

become.drunk=PERF

‘Mangta is rubbing tobacco, Bandhna *keeps [walking] here and there* as if he is being somewhat stubborn, and is quite strongly drunk.’

[Kerkeṭṭā, 1990: 24]

In fact, examples can even be found in the literature of entire sentences functioning as the semantic head of a TAM/PERSON-syntagma in quotations:

88. *u buṭha=kiyar=te=ko bay jaʔb=siʔ. iḍib*
 this old.man=DU=OBL=CNTR madness grab=PERF night

tunboʔ “kersonʔ=e la! kersonʔe la!” loʔ=na=kiyar.
 daytime marry=ACT.IRR VOC REP CNT³²=MID.IRR=DU

‘My parents have gone mad (= madness has grabbed the old man [and his wife = DU]. Day and night *they’ll keep on [saying] ‘Marry! Marry!’*’

[Kerkeṭṭā, 1990: 31]

Perhaps such constructions in Kharia are considered “inelegant” and thus to be avoided because it is impossible to construct such forms in the culturally dominant languages of the region such as Hindi and Sadri, although this topic requires more study. At any rate, such non-elicited examples from the literature make it clear that factors other than mere grammaticality also play a role in speakers’ acceptability judgements of the structures in question during interviews.

4.5 PARTS OF SPEECH IN KHARIA: A SUMMARY

In summary, with the possible exception of proforms and deictics of time and place, all non-deictic contentive morphemes in Kharia can function

³¹ *khor* is not a “verbalizer” but a v2 (6.5.2) which can attach to any semantic head compatible with its semantics of “keep on”, such as, e.g., *buli* ‘wander’. As such, it does not derive “verbs” from “nouns”.

³² *loʔ* is not a “verbalizer” but, like *khor*, also a v2 (6.5.2) and can also attach to any semantic head which is compatible with its semantics of “over and over”, such as, e.g., *chikay* ‘sneeze’. As such, it also does not derive “verbs” from “nouns”.

as the semantic head of a predicate, a referential unit, or in attributive function. More important, however, is the recognition of the fact that these units in Kharia are formed in the syntax, not in the lexicon: As all grammatical marking on these categories is enclitic, and as the semantic head may be quite complex, we must recognize the fact that these units are formed in the syntax, not in the lexicon. There are thus two structures, TAM/PERSON-syntagmas and Case-syntagmas, which may freely fill any of the three main functions of reference, predication and attribution.

The data also show that we cannot merely speak of parts of speech in "Kharia" but must differentiate between at least two dialects with subtle but important differences in this regard:

- *The Simdega dialect*: Here we find no conclusive evidence for assuming the presence of lexical categories other than "open" (contentive) and "closed", which has the two sub-classes "proforms/deictics" and "grammatical morphemes" Instead, this dialect has "syntactic parts of speech" In addition, the data suggest that any contentive morpheme may appear in both Case-syntagmas and TAM/PERSON-syntagmas when the semantic head consists of a single contentive morpheme. This holds for the reduplicated masdars as well, which are not "nouns" or "adjectives" but which can also appear in attributive, referential and even predicative function.
- *The Gumla dialect*: Here we do find evidence for a noun-verb distinction, as this has been defined here, although the data on this dialect are still rather scant. For these speakers, such complex predicates, i.e., "NPs as verbs", were generally not accepted and lexical items denoting physical or animate entities were also generally not accepted in predicative function without a light verb. A very large portion of the lexicon, however, appears to be underspecified with respect to lexical classes (cf. e.g. *biha* 'wedding; marry; married').

Nevertheless, here as well it is not inappropriate to speak of TAM/PERSON-syntagmas and Case-syntagmas—this dialect has these structural units as well. What differs is merely the acceptability of certain contentive morphemes in one class or the other as well as the potential complexity of the semantic base. Hence, as these two categories are found in both dialects, these terms will be retained throughout the remainder of this study.

4.6 “COMPOUNDS” AND “NOUN INCORPORATION”

It is general practice in Munda studies to treat two or more contentive morphemes which appear together—both of which translate as nouns in English—as a compound, although the reasons for this are seldom explicitly mentioned. However, as the large majority of the forms generally considered compounds are best considered merely juxtaposed phonological and grammatical words, at least in Kharia, we will begin in Section 4.6.1 with a discussion of what in Kharia may NOT be considered compounds but rather merely TRANSLATE as compounds into English, before discussing in Section 4.6.2 those units in Kharia which can be considered compounds, as well as a number of borderline cases in Sections 4.6.3 and 4.6.4. The chapter ends with a discussion of further “compound-like” structures in 4.7, the so-called “echo-word” constructions.

4.6.1 *Lexical juxtapositions as “compounds”*

Let us begin by briefly discussing what are NOT compounds in Kharia although they are generally considered such, presumably due to the fact that they translate into English and other languages as compounds consisting of two nominal elements. Tables 4.3 and 4.4 present a number of what Malhotra (1982: 72f.) considers to be equational and associative compounds, respectively.³³

Table 4.3: “Equational compounds”

<i>ompay kirom</i> big.river river	‘rivers’	<i>goʔjlo[ʔ]</i> field	<i>aŋkel</i> land	‘cultivated land’
<i>kheti baʔi</i> field farm	‘farms’	<i>baʔ</i> paddy	<i>rumbkuʔb</i> rice	‘grain’
<i>kinir jhaŋkoy</i> forest scrubland	‘forest, wilderness’	<i>paisa</i> money(IA)	<i>qhebuwa</i> money(Kh.)	‘wealth’

³³ I must admit that I am somewhat sceptical of a number of the interpretations of the “associative compounds” from Malhotra (1981: 73f.). For example, she translates *oʔ dura* as ‘households’, although in my corpus it appears only with reference to the physical structure of the houses, or her interpretation of *kunʔu[ʔ] hakon* ‘child blood’ = ‘kith and kin’. According to Pinnow (1965a: 93, note 36) *hakon* has the meaning ‘grandchildren’ (*Enkelkinder*) although it is no longer used as an independent morpheme by speakers I questioned. Pinnow’s analysis would seem preferable considering that *kon-* is found elsewhere in Kharia with the meaning ‘small’ and its cognate forms are found in other Austro-Asiatic languages with the meaning ‘child’.

Table 4.4: "Associative compounds"

<i>uslo?</i> earth	<i>tirib</i> cloud	'universe'	<i>iḍib</i> night	<i>meya[ʔ]</i> morning	'the whole day'
<i>ti?</i> hand	<i>kaḥa</i> leg	'body parts'	<i>konselḍu?</i> woman	<i>kopru[ʔ]</i> man	'people'
<i>kulam</i> brother	<i>kulamḍay</i> sister	'people'	<i>ḍa?</i> water	<i>pe?</i> rice	'food and drink; nutrition'
<i>oreʔj</i> bull	<i>koḥtaḥ</i> cow	'cattle'	<i>kiḥo[ʔ]</i> tiger	<i>ban[ay]</i> bear	'wild animals'
<i>lutui</i> cloth	<i>araʔj</i> flame	'everything necessary to keep one warm'	<i>beḥa</i> valley	<i>gorḍa?</i> stream	'plains'
<i>o?</i> house	<i>ḍura</i> door	'household'	<i>kunḥu[ʔ]</i> child	<i>hakon</i> blood	'kith and kin'

The status of these units as compounds is questionable: In fact, there is little reason to consider them compounds, while there is good reason to consider them simply juxtaposed elements which are closely related to one another semantically. For example, there is morphosyntactic evidence showing that possessive marking can appear either on the unit as a whole or on both of its parts individually. By way of example, let us consider the form *ayo aba* 'mother father' or simply 'parents', which would count as an "associative compound" in Malhotra's terms. The following forms of this "compound" are found in our corpus:

89. *ayo* *aba=ko* *ayo* *aba=ḍom*
 mother father=CNTR mother father=3POSS
- ayo* *aba=ḍom=kiyar* *ayo* *ro* *aba*
 mother father=3POSS=DU mother and father
 '(his/her) mother and father'

Although the possessive marker in these examples is found only attached to the final element, *aba* 'father', the fact that the two can be conjoined by *ro* 'and' is a first indication that an analysis of this unit as a compound may not be correct. Now consider the following forms, again from our corpus. Although these are different contentive morphemes, the meanings are the same and the two are merely free variants of the forms *ayo* and *aba*, respectively.

90. *mã(y) bap* ‘mother and father; parents’
ma=dom ap=dom
ap=dom ma=dom
ma=dom ro ap=dom

If these were indeed compounds, one would not expect the enclitic possessive marker to be able to appear between the two elements. Problematic is also the fact that the order of the elements is reversible: cf. *ap=dom ma=dom*. And again, the fact that the two members of the presumed compound can be conjoined by the conjunction *ro* ‘and’ seriously calls their status as a single unit into question. Of course, one could always argue that in such cases, e.g. *ayo ro aba* ‘mother and father’, the intended interpretation is not ‘parents’ but merely ‘mother and father’, but this is simply begging the question of what is meant by the term (associative) compound in the first place.

Recall also from Section 4.1 that any contentive morpheme in Kharia may be used attributively, in addition to referentially and predicatively:

91. *kuda koloŋ daru* ‘a millet bread tree’ (from a children’s story)
 millet bread tree

As we shall see in Section 5.2, in cases such as these, the genitive may also freely appear, apparently with no semantic differences, instead of the mere juxtaposition:

92. *kuda koloŋ =ya? daru* ‘a millet bread tree’
 millet bread=GEN tree

This clearly favors the interpretation of *kuda koloŋ* and *daru* as separate syntactic units in juxtaposition to one another, and the same relationship then also holds between *kuda* and *koloŋ* as well.

Finally, as Rehberg (2003: 31f.) notes with respect to two units which would have to be classified as “attributive compounds” in Malhotra’s terminology, such as *duŋduŋ kaŋoŋ* [eel fish] ‘eel’ and *ɖakay rani* [Dakay queen] ‘Queen Dakay’, there is no positive evidence for considering two units such as these to form a single phonological unit. Recall in this respect also *manus jati* [human caste] ‘humanity’ in Diagram 2.19 in Chapter 2, which shows that both units have the same underlying LH pitch pattern within the same frequency range. Thus, there is no phonological evidence for assuming that these two elements form a tightly knit unit.

Thus, in the lack of any positive evidence for considering these units compounds, and since there is considerable evidence against doing so, all examples of this type are considered here to be merely juxtapositions, not compounds.

4.6.2 *Genuine compounds*

There are two types of units in Kharia for which the status “compound” appears appropriate:

- *Structural considerations*: Units in which one of the components appears in the so-called “compound form”, or units in which one component should appear in its reduplicated form (the masdar, 4.3.2) as an independent phonological word but does not;
- *Semantic considerations*: Units which together have a meaning which is not equal to the sum of the individual meanings.

Structural considerations: the compound form and reduplication

Recall from Section 4.3.1 that Kharia requires phonological words to be bisyllabic, with very few exceptions. Thus, virtually all contentive morphemes which can stand alone as a phonological word contain at least two syllables, such as *uslo?* ‘land’, *lo?kha* ‘earth, soil’, *siŋkoy* ‘chicken’ and *soreŋ* ‘stone’. On the other hand, there is the so-called compound form, where this same element—semantically speaking—contains only one syllable, compare *-lo?*, *-siŋ* and *-sor* with the forms just mentioned. This form is never found as a syntactic atom of the clause but only in conjunction with another contentive morpheme. These units may thus safely be considered compounds. Table 4.5 gives a few examples.

Table 4.5: Compounds in Kharia (adapted from the examples in Malhotra, 1982:74f.)

<i>ba?-lo?</i> paddy-land	‘paddy land’	<i>ja?-lo?</i> grass-land	‘meadow’
<i>dhuŋi-lo?</i> dust-land	‘desert’	<i>kunŋu[ʔ]-siŋ</i> child-fowl	‘chicken’ ³⁴
<i>si-lo?</i> plow-land	‘plow (v.)’	<i>pa?-sor</i> break-stone	‘grinding slab’

³⁴ Most likely a chick is meant here, i.e. a ‘baby chicken’

Another type of compound can be identified by the fact that an underlying monosyllabic morpheme, which as a phonological word would obligatorily reduplicate, is found in its non-reduplicated form. This may also be taken as conclusive evidence for analyzing such forms as compounds. Cf. again *paʔ-sor* 'grinding slab' (vs. *paʔ-paʔ soreŋ* 'broken stone') as well as the following forms:

93. <i>iʔj</i> 'defecate'	<i>iʔj-thaŋ</i> 'defecate (of cows)'	(< <i>*taŋ</i> 'cow') ³⁵
<i>muʔ</i> 'emerge, come out'	<i>muʔ-ɖaʔ</i> 'source; origin(ate)'	(cf. <i>ɖaʔ</i> 'water') ³⁶
<i>muʔ</i> 'emerge, come out'	<i>muʔ-rel</i> 'rise (of the moon, stars)'	(< <i>*ler</i> < <i>leraŋ</i> 'moon, month') ³⁷
<i>muʔ</i> 'emerge, come out'	<i>muʔ-siŋ</i> 'rise (of the sun)'	(< <i>siŋkom</i> 'star') ³⁸
<i>soʔ</i> 'fasten, tie (hair)'	<i>soʔ-lui</i> 'tie up a girl's hair; girl's hair'	(< <i>ului</i> 'hair')
<i>ur</i> 'get released; shake oneself free'	<i>ur-boʔ</i> 'shake the head'	(< <i>*bok</i> < <i>bokoʔb</i> 'head') ³⁹

It is important to note here that this is not a productive process in modern Kharia, i.e., these forms may only be considered compounds from a historical perspective. From a synchronic perspective they are best considered contentive morphemes in their own right.

³⁵ While not a free morpheme nor productive, *-t(h)aŋ* appears in a number of words related to cows: *koŋtaŋ* 'cow', *ɖintaŋ* 'stall', *giniŋ-taŋ* 'bride price' (formerly paid in cattle).

³⁶ As opposed to *muʔ-muʔ ɖaʔ*, with the reduplicated masdar and meaning 'water which has emerged'. Note, however, that one of the two southern speakers I worked with uses the form *muʔ-muʔ ɖaʔ* with the meaning 'origin', which however was rejected by other speakers and is clearly a dialectal form.

³⁷ Cf. Pinnow (1959: 77, §58). *-rel* is also a common formative in the Kharia names of the months of the year, such as *magh-rel* 'January' from the Indo-Aryan loan *māgh* 'the month of Magh'.

³⁸ The contentive morpheme *siŋkom* in Kharia seems to have undergone a semantic shift since this type of incorporation was productive. However, cognate contentive morphemes in related languages still retain this meaning. Cf. *siŋi* with the meaning 'sun' in the related South Munda language Remo/Bonda (Bhattacharya, 1968: 127; Fernandez, 1967: 154), as well *siŋi* in the North Munda language Ho (Deeney, 1975: 319, with the additional meaning 'daylight').

³⁹ Underlying /k/ and /g/ in the coda are realized as [ʔ]. From Biligiri (1965: 205).

Semantic considerations

Only a few forms have been noted in which the semantics of a particular unit do not equal the sum of its parts but must be learned as a whole, although its individual members constitute full words, generally one of the small number of monosyllabic contentive morphemes which do not reduplicate as *masdar*.⁴⁰ Consider the forms shown in Table 4.6. This is a highly problematic criterion when it is not also supported by structural criteria, however, as it is not always clear where to draw the line between predictable and non-predictable semantics.

Table 4.6: Semantically unpredictable compounds
(examples from Malhotra, 1982: 74ff.)

<i>diyom-jar</i> lamp-bone	'knee bone', lit. 'lamp-shaped bone'	<i>kaḍoŋ-mo'd</i> fish-eye ⁴¹	'name of tree (whose leaves are shaped like fish eyes)'
<i>goŋŋiŋ-o?</i> cook-house	'kitchen'	<i>borol-ḍa?</i> live-water	'fresh water'

4.6.2.1 *(Semi-)Productive compounding*

There are also a few more-or-less productive types of genuine compounding in Kharia. Their status as compounds is shown by the fact that (underlying) monosyllabic contentive morphemes appearing within them do not reduplicate, hence the two units must form a phonological word together. The only three types which have so far been noted are the following.

- *jaw-*: 'until', combines with the contentive morpheme. E.g., *jaw-ḍel* 'until [someone] comes', e.g., *ho=kaŋ=a? jaw-ḍel* [that=SG.HUM=GEN until-come] 'until s/he comes'

Note that these two units must form a compound, as the contentive morpheme is not reduplicated if it is monosyllabic, i.e., it is not a *masdar*. This is an entirely productive compounding process. It is certainly related to the adposition *jou / jaw* 'up to', which however is postposed (5.3).

⁴⁰ For reasons of space, I will not repeat here those forms which have slightly irregular semantics but which can already be considered compounds for purely structural reasons, such as *pa?-sor* [break-stone] 'grinding slab', discussed above.

⁴¹ My gloss, as Malhotra gives none here.

- *-mon* ‘mind; wanting to’, similar to the *jaw*-construction in that the lexical unit is not reduplicated. Only marginally productive. E.g., *aw-mon* [stay-mind] ‘wanting to stay’
- *-ḍay*, with the intervocalic allomorph *-ṛay*, may be considered the compound form of *kandḍay* ‘woman’. This construction is productively used to denote females: *kulam* ‘brother’ (more properly: ‘sibling’) vs. *kulam-ḍay* ‘sister’, *sou* ‘husband’ (more properly: ‘spouse’) vs. *sou-ṛay* ‘wife’. However, for some speakers, especially of the southern dialects, *ḍay* is an independent contentive morpheme.

94. *khariya ḍay=ki=ko buʔjhi guḍ=siʔ=ki.*
 Kharia woman=PL=CNTR understand C.TEL=PERF=PL
 ‘The Kharia women understood [the situation].’ [MT, 1:202]

At present, we know of no other types of genuine compounds from a purely synchronic perspective in Kharia.

4.6.3 Pseudo-compounds and “noun incorporation”

Compounds (from a historical viewpoint) of the type discussed in Section 4.6.2 above are in fact quite common, although this is no longer a productive process. These are generally found in predicative function and are then referred to as “noun incorporation” in Munda studies, a process which is still productive in some South Munda languages.

The “incorporated nominal” often appears to be the subject or object of the predicate, as in the following example:

95. *si* ‘plow’ (somewhat archaic) *si-loʔ* ‘plow (v.)’ (cf. *loʔkha* ‘earth’)

Nevertheless, compounds such as *si-loʔ* are fully opaque to syntax: Even though it would appear that the object has been incorporated into the predicate in (95), as the following example shows, the “incorporated object” is not the object of the clause but rather *goʔjloʔ=te* ‘the field’:

96. *hoḍom=ki=yaʔ goʔjloʔ=te si-loʔ=na ayiʔj*
 other=PL=GEN rice.field=OBL plow-earth=INF QUAL.PRS
laʔ madet remaʔ=te=ki.
 then help call.for=ACT.PRS=PL
 the rice field of the others must be plowed (= is to plow), then they call for help.’ [AK 5,13-14]

Similarly, with respect to what would appear to be a “locative argument”:

97. *tomleŋ*=te su⁷b
 milk=OBL reach.hand.into
ɖaʔ=yoʔ ‘He *reached his hand* into the [bucket of] milk.’
 water=ACT.PST

As has long been noted (e.g., Pinnow, 1960: 97), the apparent “Verb-Object” order of these compounds would seem to be a sign of their high antiquity, deriving from an earlier stage in which the Munda languages, like the Mon-Khmer languages of Southeast Asia today, had the underlying constituent order S-V-O. Nonetheless, from a purely synchronic viewpoint, these forms are merely contentive morphemes in their own right in Kharia, not compounds.

There is another, highly similar construction in Kharia with this same “Verb-Object” order, which would also seem to indicate that these forms are quite old:

- | | |
|--|--|
| 98. <i>ajo²ɖ</i> ‘dry up’ (ITR) | <i>ajo²ɖ ɖaʔ</i> ‘dry up (cf. <i>ɖaʔ</i> ‘water’)
(of water)(ITR)’ |
| <i>u²ɖ</i> ‘drink’ | <i>u²ɖ ɖaʔ</i> ‘drink (not only water!)’ |
| <i>su⁷b</i> ‘reach the hand into’ | <i>su⁷b ɖaʔ</i> ‘reach the hand into a fluid’ |

Similar to the (historical) compounds discussed in Section 4.6.2 above, this is not a productive process and new forms cannot freely be constructed:

99. *suluy* ‘warm up’ (ITR) **suluy ɖaʔ*

Also similar to the compounds discussed above, these forms are most commonly found in predicative function:

100. *ɖaʔ* ajo²ɖ ɖaʔ=ki.
 water dry.up water=MD.PST
 ‘The water *dried up*.’
101. *caha* u²ɖ ɖaʔ=yoʔ.
 tea drink water=ACT.PST
 ‘S/he *drank* tea.’

Nonetheless, despite all similarities with compounds, these forms are neither compounds nor derived forms but merely juxtaposed contentive morphemes, where the final morpheme is one of the few forms in the language which, although monosyllabic, are nevertheless free-forms. As the following example shows, the two elements of this apparent compound may be separated from one another, e.g., by enclitic pragmatic markers, which strongly calls their status as compounds or lexemes into question:

102. *qa?* *qjo'd=ga* *qa?=ki.* 'The water dried up.'
 water dry.up=FOC water=MID.PST

While some speakers found this last example to be somewhat odd, although not ungrammatical, they had no problems with the following pair:

103. *urum* *qa?=ki.* 'S/he sweat.'
 warm water=MID.PST

104. *urum=ga* *qa?=ki.* 'S/he sweat.'
 warm=FOC water=MID.PST

This was different for the (historical) compounds discussed in the last section, where the insertion of an enclitic marker was consistently rejected:

105. *ol-day=o?* **ol=ga* *day=o?*
 take-woman=ACT.PST
 'He married.'

106. *boŋ-sor=ki.* **boŋ=ga* *sor=ki.*
 finish-stone=MID.PST
 'It petrified.'

The group of complex forms in (98) will be referred to here as "pseudo-compounds": Although the elements do not form a tight syntactic bond and can be separated from one another at least by enclitic pragmatic markers, their semantics are nevertheless somewhat idiosyncratic. For example, *qa?*, which otherwise translates as 'water', refers in these units merely to any fluid, whereas with *urum qa?* 'sweat', the resultant meaning is not equal to the sum of its components, i.e., 'sweat' ≠ 'become

warm' + 'water' Rather, these units are more akin to idiomatic expressions with a fixed, somewhat unpredictable meaning.

Below is a list of all cases of such non-productive "incorporation", both compounds and pseudo-compounds, that have been identified so far. In these examples, "*" denotes a reconstructed form. Many of these units were already discussed with respect to their status as (historical) compounds in Section 4.6.2 above but are repeated here for convenience.

The speakers we worked with could often not further analyze the forms presented here as historical compounds, although they had no problems analyzing the pseudo-compounds. The historical compounds are thus simply (derived) contentive morphemes for these speakers. In these cases, the analysis here is our own or, in the case of *ur-bo?*, from Biligiri (1965).⁴²

107. Final element as "subject"—Historical compounds:

<i>iʔj</i> 'defecate'	<i>iʔj-thaŋ</i> 'defecate	(< * <i>taŋ</i> 'cow')
		(of cows)'
<i>mu?</i> 'emerge, come out'	<i>mu?-da?</i> 'originate; source' ⁴³	
<i>mu?</i> 'emerge, come out'	<i>mu?-rel</i> 'rise (of the moon, stars)'	(< * <i>ler</i> < <i>leran</i> 'moon, month')
<i>mu?</i> 'emerge, come out'	<i>mu?-siŋ</i> 'rise (of the sun)'	(< <i>siŋkom</i> 'star')
<i>tu</i> 'DEPARTIVE', apparently once used as a lexical morpheme (6.5.7)	<i>tu-da?</i> 'carry off (of a river)'	
<i>ur</i> 'get released; shake oneself free'	<i>ur-bo?</i> 'shake the head'	(< * <i>bok</i> < <i>bokoʔb</i> 'head')

Pseudo-compounds:

<i>ajoʔd</i> 'dry up' (ITR)	<i>ajoʔd da?</i> 'dry up (of water)(ITR)'	(cf. <i>da?</i> 'water')
<i>boŋ</i> 'come to an end'	<i>boŋ da?</i> 'come to an end, be depleted (of water)'	
<i>ulu?</i> 'boil' (ITR)	<i>ulu? da?</i> 'boil (of water)' (ITR)	

⁴² In fact, they occasionally outrightly rejected my own and Biligiri's analyses. Interestingly, however, they had no problems in analyzing the *-sor* in *boŋ-sor* as related to *soreŋ* 'stone', undoubtedly due to the semantics.

⁴³ See note 74 above.

108. Final element as “object”—Historical compounds:

<i>guʔj</i> ‘wash’	<i>gʊjun</i> ‘wash the feet’	(< * <i>j(h)un</i> ‘foot’) ⁴⁴
<i>guʔj</i> ‘wash’	<i>guʔthe, guʔthe</i>	(< <i>tiʔ</i> ‘hand’) ⁴⁵
	‘wash the hands’	
<i>ol</i> ‘take’	<i>ol-ɖay</i> ‘marry’	(< <i>ɖay</i> ‘woman’)
<i>si</i> ‘plow’	<i>si-loʔ</i> ‘plow (v.)’	(< <i>loʔkha</i> ‘earth’)
<i>soʔ</i> ‘fasten, tie (hair)’	<i>soʔ-lui</i> ‘tie up a girl’s hair; girl’s hair’	(< <i>ului</i> ‘hair’)

109. Pseudo-compound:

<i>uʔd</i> ‘drink’	<i>uʔd ɖaʔ</i> ‘drink (liquid)’
--------------------	---------------------------------

110. Other functions—Historical compound:

<i>boŋ</i> ‘come to an end, finish’ (ITR)	<i>boŋ-sor</i> ‘petrify’	(< <i>soreŋ</i> ‘stone, rock’)
---	--------------------------	--------------------------------

Pseudo-compounds:

<i>betoʔd</i> ‘become hungry’	<i>betoʔd ɖaʔ</i> ‘become thirsty’
<i>guʔj</i> ‘wash (clothes, etc.)’	<i>guʔj ɖaʔ</i> ‘soak in water; washing water’
<i>suʔb</i> ‘reach the hand into s.th.’	<i>suʔb ɖaʔ</i> ‘reach the hand into a fluid’
<i>urum</i> ‘become warm’	<i>urum ɖaʔ</i> ‘sweat’(n./v.) ⁴⁶

Finally, there are a few expressions which appear to have derived from this construction but whose “verbal” component is no longer found as a contentive morpheme:

111. <i>kul-ɖaʔ</i> ‘get/have a fever’	* <i>kul ɖaʔ</i> ‘water’
--	--------------------------

4.6.4 *Semi-productive incorporation*

Some speakers will accept the incorporation of non-specific and non-referential objects into the predicate. The incorporated element in this construction appears directly after the reciprocal marker *kol*, if present,

⁴⁴ As Pinnow (1959: 169, §365; 202, §61) notes, this is undoubtedly from an old compound, consisting of *guʔj* ‘wash’ and **jun* ‘foot’. Cf. also *goʔdʒun* ‘path’ and *toʔjʒun* ‘stuck the toe’, both of which contain *j(h)un* and are related to the feet.

⁴⁵ Cf. Pinnow (1965b: 66, note 52).

⁴⁶ Biligiri (1965: 295) also analyzes this as I have, but only gives *urumɖaʔ* as a “noun”.

and before the remaining semantic base. All speakers who accepted these forms, however, preferred forms in which the object was not incorporated. All forms given here were elicited in interviews. Only minor, optional phonetic changes, such as the loss of /ʔ/ in the coda, occur.

112. Acceptable (to some) Preferred
kol ula? likha=ki=kiyar *ula? kol likha=ki=kiyar*
 REC letter write=MID.PST=DU
 ‘they wrote each other letters’
113. *kol qa uq=ki=kiyar* *qa? kol uq=ki=kiyar*
 REC water drink=MID.PST=DU
 ‘they gave each other water to drink’

This type of incorporation is limited to certain commonly occurring combinations, such as *letters* and *writing*, *water* and *drinking*, etc., but is not possible in other combinations:

114. **kol tomliŋ uq=ki=kiyar* ‘they gave each other milk to drink’
 REC milk drink=MID.PST=DU

It is as yet unclear to what extent this involves compounding, as the data for this (very rare) construction are still rather scant. What is clear, at any rate, is that *kol* and the following, incorporated element do not form a tighter bond, as at least enclitic pragmatic markers can intervene between the two.

115. *kol=ga ula? likha=ki=kiyar.* ‘They wrote each other letters.’
 REC=FOC letter write=MID.PST=DU

To what extent the incorporated element forms a compound with the following unit is unclear.

4.7 “ECHO-WORDS”

Kharia has a construction which is quite similar semantically to what is often termed the “echo-word” formation in South Asian linguistics, in which a lexical stem is followed by an element which has no independent meaning but which modifies the meaning of the first element, generally

indicating something akin to the English “etc.” The “echo-word” construction in Kharia is actually a collection of forms from at least three different sources, which will now be discussed separately.

The first consists of forms in which the second element seems to have once had an independent meaning which is no longer commonly known. In all cases it appears that the original meaning of this element was very close to that of the first element as in forms such as the following, of which both of the respective elements are still in common use: *ompay khiro* ‘big.river river’ = ‘rivers’, *ba? rumku?b* ‘unhusked.rice husked.rice’ = ‘rice’ or *paisa dhebua* ‘money money’ = ‘wealth’ (adapted from Malhotra, 1982: 72f. and discussed above in Section 4.6.1).

Consider now *kinir jhaṅkoy* / *kinir jhaṅkor*, which Malhotra (1982: 73) translates as ‘forest, wilderness’ and analyzes as a compound consisting of *kinir* ‘forest’ and *jhaṅkoy* (also realized as *jhaṅkor*, JP) ‘scrubland’. Some speakers were not familiar with this second term in isolation, although it would seem to have lost its independent meaning only recently. It is also attested in Roy & Roy (1937: 165) with the form *jhaṅkor* and the meaning ‘forest’ or ‘*sarna* (holy forest)’ and is perhaps related to *jaṅkor* / *jaṅkoy* ‘spring (festival) or ‘festival of flowers’

We can probably also add other expressions to this category, such as *khe?d jol* ‘bite etc.’ from *khe?d* and the echo-word *jol* (?). This analysis is supported by the fact that when these forms are used predicatively, the first element may also be partially finite, as in the following example, suggesting that both elements once had independent meanings and that this is not a compound form:

116. *khariya dhirom dhirom buli=ga armaray=ga ki?o?, buwan,*
 Kharia slowly REP wander=FOC hesitate=FOC tiger snake
bichi buṅ khe?d ɖuṅ =ga jol ɖuṅ =ga ɖel=ki=may.
 scorpion INST bite PASS=FOC ECHO PASS=FOC come=MID.PST=3PL
 ‘The Kharia came, wandering slowly, stopping, *being bitten*, etc.,
 by tigers, snakes and scorpions.’ [MT, 1:13]

Similarly we find from the same speaker the expression *til=o? raph=o?[=ki]* [bury=ACT.PST ECHO=ACT.PST=PL] ‘they buried’, where the second contentive morpheme, *rab* (> *raph* before =o?) with the meaning ‘bury’, is found in Roy & Roy (1937) and Pinnow (1965a) but is now unknown to most speakers.

The second group is quite large and consists of loan words from Sadri. Here we present just a few examples: *lare bhire* / *larai bhirai* 'fighting, etc.', *kado walo* 'mud, etc.', *culha cawka* 'stove, etc.' Often, the second component is still in current use with approximately the same meaning as the first element: *main marjad* 'honour, etc.', *khøj puchar* 'investigation, etc.'

Finally, there is a productive construction in Kharia with a type of repetition in which one element, usually the second, is a repetition of the first, but phonologically somewhat different. The resultant meaning is usually that of a durative or gradual telic action, although not always. Also, there is some variation from one speaker to another with respect to the form of the repeated element, so that no general rules can be given with respect to the structure of the repeated unit other than that the repeated unit usually substitutes /i/ for at least one of the vowels of the underlying form, often adding a final -a. Consider the forms shown in (117).

- | | |
|---|--|
| 117. <i>bi²d</i> 'pour out; sow' | <i>bi²d bi²a</i> 'pour out completely;
sow everything' |
| <i>cipa</i> 'squeeze out' | <i>cipa cipi</i> 'squeeze out
(HJPA, 154, d:14; 157, fn. d:14) |
| <i>col</i> 'go' | <i>col cila</i> / <i>col ciya</i> 'go away' |
| <i>d²el</i> 'come' | <i>d²el d²ila</i> 'arrive, come all the way' |
| <i>d²ubay</i> 'press down,
deplete' | <i>d²ubi d²ubay, d²ubay d²ubi</i>
'completely deplete' |
| <i>dho?</i> 'grab, take' | <i>dho? dhiga</i> 'take, grab
(e.g., an electric wire which one is
not supposed to touch; steal s.th.)' |
| <i>gam</i> 'say' | <i>gam gim</i> 'talk s.th. out, say all there is
to say; speak when one should not;
use improper speech' |
| <i>gupa</i> 'watch over
(cattle, etc.)' | <i>gupa gupi</i> 'finish tending' |
| <i>kayom</i> 'talk' | <i>kayom kiyim</i> 'talk s.th. out' |
| <i>kuda</i> 'hurry' | <i>kuda kudi</i> 'hurry, scramble' |
| <i>le²j</i> 'curse' | <i>le²j lija</i> / <i>le²j liya</i> 'curse s.o. badly,
entirely' |
| <i>leme²d</i> 'lie down; sleep' | <i>leme²d limi²d</i> 'fall asleep' |
| <i>likha</i> 'write' | <i>likha likhi</i> 'write out' |

<i>melay</i> ‘leave’	<i>meli melay</i> ‘leave behind’
<i>no?</i> ‘eat’	<i>no? jiga / no? juya?</i> ‘eat up’
<i>poṛhe</i> ‘read’	<i>poṛhe piṛhi</i> ‘read completely’
<i>sangoʔd</i> ‘walk; close’	<i>sangoʔd sinjiʔd</i> ‘walk all the way; close completely’
<i>soŋ</i> ‘buy’	<i>soŋ siŋa</i> ‘sell off’
<i>uʔd</i> ‘drink’	<i>uđ iđa?</i> ‘drink up’
<i>uwa?</i> ‘wash; bathe’	<i>uwa? uwi?</i> ‘finish washing up / bathing’

There are also a few forms in the data which appear to belong to this construction although the corresponding simplex is not attested, such as *ceŋga ceŋgi* ‘constantly change one’s mind’

Note also that with bisyllabic contentive morphemes the repeated element can appear before the underlying form itself, as with e.g. *ḡubi ḡubay* and *meli melay*. The following presents two examples of this construction from the corpus.

118. *oʔ=te* *ap=ḡom* *teinko* *arkhi* *un=sig=e*
house=OBL father=3POSS a.little liquor place=PERF=ACT.IRR

ho=jeʔ=ko *ho=kaṛ* *capu* *karbaʔd=kon*
that=SG.NHUM=CNTR that=SG.HUM rummage ECHO=SEQ

uʔd *iđaʔ=na* *laʔ=ki.*
drink ECHO=INF IPFV=MID.PST

‘If his father has placed some liquor in the house, then he would rummage through everything [until he found it] and then *drink it all up.*’
[RD, 2:25]

119. *puŋji=te* *ḡubi* *ḡubay=kon* *ap=ḡom=te* *ođoʔ=ga*
funds=OBL ECHO deplete=SEQ father=3POSS=OBL more=FOC

keciya *bor=na* *laʔ=ki.*
money ask.for=INF IPFV=MID.PST

‘When he *had depleted* his funds he would ask his father for more money.’
[RD, 2:37]

As the following example shows, the “echo-word” may also be separated from the primary lexical morpheme by *no* ‘and; or; CMPL’ when

used as an “imperfective converb” (6.6.2.4), again showing that this complex unit is not a compound:

120. *musniŋ ho=ki jhaŋi kuru? hakon=ki=te somtay=oʔ=ki*
 one.day that=PL all child ECHO=PL=OBL collect=ACT.PST=PL
- ro kuda no kudi raylogaḥ tay yar kan=oʔ=ki.*
 and hurry Cmpl ECHO Railogarh ABL flee CONT=ACT.PST=PL
 ‘One day they gathered up all the children and fled quickly (= *hur-*
rying) from Railogarh.’ [MT, 1:105]

THE CASE-SYNTAGMA (“NPs”)

5.1 GENERAL INTRODUCTION

Although it was shown in Chapter 4 that any syntactically well-formed semantic base—whether simple or complex—may function as the semantic base of both a TAM/PERSON-syntagma or a Case-syntagma, for ease of presentation and reference those categories which are generally found in the semantic base of a Case-syntagma, such as number marking, demonstratives, etc., will be discussed in this chapter, although it should be stressed here once again that it is only case marking which defines a Case-syntagma as such. That is, the general structure of a Case-syntagma is as in Table 5.1, where “X” is the possibly complex semantic base of the Case-syntagma:

Table 5.1: The structure of a Case-syntagma in Kharia

X=CASE

“Case” in this sense refers to the direct case (unmarked), the oblique case (=te), and adpositions but not to the genitive, since as we saw in Section 4.4, the genitive can appear in the semantic base of both TAM/PERSON- and Case-syntagmas.¹

The maximal structure of a Case-syntagma can be summarized schematically as in Table 5.2, where the maximal possible structure of the semantic base (“X” in Table 5.1) has been expanded.

Table 5.2: A schematic overview of a Case-syntagma

GENITIVE DETERMINER DEM QUANT CLASS GENITIVE DETERMINER LEXEME(S)=POSS=NUM/
HON=CASE

¹ Note that this is slightly different from the use of the term “case” in Section 5.2, where “case” is distinguished from “adposition” and includes the genitive. As unfortunate as this double usage of the term “case” is, it seemed preferable to coining a new term for either use.

1. *nāw jan beʔt=dɔm=kiyar* ‘their nine sons (HON)’
 nine CLASS SON=3POSS=HON
2. *ho=kiyar=ya? carpaʔ* ‘their cleverness’
 that=DU=GEN cleverness
3. *ho pāc lebu=ki* ‘those five persons’ [Biligiri, 1965: 77]
 that five person=PL

In addition, there may be a relative-clause-like modifier which generally precedes the entire structure shown in Table 5.2 but which may also follow the demonstrative. These structures will be dealt with in more detail in Sections 5.8 and 7.6.

There is no compelling evidence for the presence of “nouns” and “adjectives” in (Simdega) Kharia. Instead, contentive morphemes can simply be regarded as juxtaposed units, each comprising a phonological word, with the semantically “most central” unit appearing to the far-right. This is denoted in Table 5.2 by the term “Lexeme(s)”, denoting one or more contentive morphemes in juxtaposition.

4. *ho rusuŋ oʔ* ‘that red house’
 that red house
5. *moŋ maha daru=ya? tuta=te* ‘at the bottom of a big tree’
 one big tree=GEN bottom=OBL

The same is also true for what would appear to be compounds, at least from an English-speaking perspective. As there is no evidence that these two or more units should be considered compounds (cf. §4.6), they are simply considered juxtaposed contentive morphemes.

6. *beʔi konʦuʔ ho laro? daru*
 girl child that date.palm tree
 ‘girl’ ‘that date-palm tree’ [Biligiri, 1965: 77]
7. *kinir kontheʦ* 8. *raksin kãʦayboʔ*
 forest bird witch old.woman
 ‘birds of the forest’ ‘the old witch’

From a children’s story:

9. *kuda koloŋ daru bo? =te* ‘at the place where the millet bread tree was’
 millet bread tree place=OBL

All of the “lexical” morphemes given in Table 5.2 above are actually *masdars* (6.6.2.1). With few exceptions, monosyllabic morphemes which appear anywhere in the semantic base of a Case-syntagma generally reduplicate, as with *su-su* in the following example, whereas polysyllabic morphemes do not:

10. *beŋi =ya? su-su lutui =te*
 girl=GEN put.ON-RDP clothes=OBL
 ‘the clothes (OBJ) that the girl had been wearing’

Enclitic focal markers (restrictive (=ga), contrastive (=ko) and additive (=jo) focus) generally attach to the last element of a Case-syntagma:

11. *am =a? haŋ =no =m =te =jo* ‘your bones (OBJ) as well’
 2SG=GEN bone=2POSS=2SG=OBL=ADD

They can, however, mark any constituent of the semantic head, with the exception of the demonstratives.²

12. *ho rusuŋ =ga o?* ‘that RED house’
 that red=FOC house

13. *ek =ga t̄ho (=ga)* ‘one (and only one)’
 one=FOC CLASS=FOC

14. **ho =ga rusuŋ o?*
 that=FOC red house

Although not a hard-and-fast rule, there is a clear preference for possessive genitive determiners to appear to the far left in the semantic base

² However, in the presence of the “classifier” *mene* (5.9.2) the clitics may directly follow the demonstrative.

with the second genitive determiner position reserved for the material something is made of or where it is from:

15. *adji=ya?* *moŋ dhaŋgar=qom=te gam=o?*...
 ANAPH=GEN one servant=3POSS=OBL say=ACT.PST
 ‘He said to one of *his* servants... [Malhotra, 1982: 127]

16. *moŋ kinir=a?* *jantu*
 one jungle=GEN animal
 ‘an animal *of the jungle* (i.e., a wild animal)’

The markers for inalienable possession, number and case to the right in Table 5.2 are all enclitic. For example, if the lexical base is not explicitly mentioned (e.g., if it is known), these markers simply attach to the right-most element of the semantic head, regardless of its status, whether this is a contentive morpheme or a genitive-marked unit. Thus, in the following example, the contentive morpheme *lebu* ‘person’ may be omitted when it is clear from context.

17. *munu?siŋ rochob=a?**ki=ko* (cf. *munu?siŋ rochob=a?* *lebu=ki=ko*)
 east side=GEN=PL=CNTR east side=GEN person=PL=CNTR
 ‘the easterners’ (FOCUSED) ‘the *people* of the east’ (FOCUSED)

In fact, if there is no further information within the semantic head, as this is all known from context, these markers then attach directly to the demonstratives, which then serve as “pronominals” of the 3rd person, e.g. *ho=ki=te* ‘them’ from *ho* ‘that’

In the following, the status of gender / size and animacy (5.1.1) and definiteness (5.1.2) as grammatical categories in Kharia is discussed. The categories represented in Table 5.2 are then discussed individually. As case marking and adpositions define a Case-syntagma as such, these two categories are dealt with in Sections 5.2 and 5.3. Strictly speaking, the remaining categories can belong to the semantic base of either a TAM/PERSON- or a Case-syntagma, however, for ease of presentation and reference, they too will be discussed in this chapter, beginning with the enclitic markers of number (5.4) and inalienable possession (5.5). Following this the remaining sections deal with proforms (5.6), interrogatives and indefinites (5.7), the demonstratives (5.8), quantifiers and classifiers (5.9) and modification (5.10), which includes both morphologically sim-

ple attribution or "adjectives" (5.10.1) and clausal adjuncts or "adverbials" (5.10.2).

5.1.1 *Gender / size and animacy as grammatical categories*

There is no grammatical gender in Kharia. There are, however, several means of expressing whether a male or female person or animal is being referred to. One means is by indicating sex lexically: *kōpuṛu?* 'man', *konseldu?* 'woman', *kongher* 'boy', *konsel* 'girl', *kokro sinḡkoy* 'cock', *kitur sinḡkoy* 'hen' (the last two from Malhotra, 1982: 66).

Another common means is through the morpheme *-ḡay* 'woman', with the intervocalic allomorph *-ṛay*, e.g. *saw/sou* 'husband', *saw-ṛay/sou-ṛay* 'wife', *kulam* 'brother', *kulam-ḡay* 'sister', etc. (Malhotra, 1982: 67, cf. also 4.6). For some speakers, *ḡay* 'woman' is a free contentive morpheme, for others it is restricted to compounds of this type. When there is no direct evidence for considering it a free contentive morpheme for a particular speaker, such as the ability to take a genitive attribute, it will be considered a bound morpheme and attached to the preceding element graphically through the symbol "-"³

A number of lexical morphemes which are specified for gender have been borrowed from Indo-Aryan, such as *koṛhi* 'lazy' (m.), *koṛhni* 'lazy' (f.). In the large majority of these forms, however, the masculine form ends in /a/ and the feminine in /i/, e.g. *beṭa* 'son', *beṭi* 'daughter', *budḡha* 'old man', *budḡhi* 'old woman'. These same two endings can also denote large/neutral vs. small entities. Again, this use has been borrowed from Indo-Aryan, such as *ḡonga* 'boat' vs. *ḡongi* 'small boat' (cf. Biligiri, 1965: 144f), *ceṇna* 'chicken, chick (larger)', *ceṇni* 'small chick', *ḡhaṇṭa* 'bell', *ḡhaṇṭi* 'small bell', etc.

This use is not productive and is restricted to Indo-Aryan borrowings. It also has no further morphosyntactic ramifications, especially since the "pronominal" system distinguishes neither natural gender nor size (5.6).

With respect to animacy as a grammatical category: Biligiri (1965) assumes a [\pm animate] distinction in the lexicon, based primarily on the fact that animate "nouns" can occur in all three numbers while inanimates do not mark for number (Biligiri, 1965: 36). There are, however, numerous

³ Its use as a free morpheme appears to be restricted to the more southerly Kharia-speaking areas.

counter-examples in our corpus. Here two examples from the first story in our corpus:

18. *kahani lebu=ki khoṭi=ki=te kayom=ta=ki.*
 story person=PL village.section=PL=OBL speak=MID.PRS=PL
u=ghay ayiʔj.
 this=way QUAL.PRS
 ‘The people tell [this] story *in the villages*. It goes like this.’
 [AK, 1:2]

19. *ro beʔt=d̥om=ki=jo umay juṇ=oʔ no i jhāut heke*
 and son=3POSS=PL=ADD NEG.3PL ask=ACT.PST Cmpl what animal QUAL.PRS
lekin kaʔ kom=ki dhoʔ=ke muʔ goʔd̥=ki=may.
 but bow arrow=PL grab=SEQ emerge C:TEL=MID.PST=3PL
 ‘And his sons also didn’t ask which animal it is, but, having taken
 their *bows and arrows*, they went off.’
 [AK, 1:11]

As the following example from Pinnow (1965a: 40) shows, this does not appear to be a recent innovation, although the use of plural marking with inanimates may be increasing, considering typical paths of grammaticalization (e.g. Lehmann, 1995):

20. *jinyray ho=kaṭ=aʔ noʔ=na=? ghaḍ diyo=ga poṛob=ki*
 porcupine that=SG.HUM=GEN eat=INF=GEN PURP daily=FOC yam=PL
ol=na laʔ=ki.
 bring=INF IPFV=MID.PST
 ‘The porcupine daily brought her *yams* to eat (= for her eating).’

Contentive morphemes with inanimate reference do show a stronger tendency to remain unmarked for number even when referring to plural entities than do those referring to animates:

21. *tay moṇ moṇ bhai=ki=yaʔ jimi dhoʔd̥=ta dhoʔd̥ta ho=kaṭ*
 then one REP brother=PL=GEN name take=CVB REP that=SG.HUM
ḥhisa=na māṛe=yoʔ.
 call.out=INF start=ACT.PST
 ‘Then, taking *the names* of the brothers one by one, he began to call
 out [to them].’
 [AK, 1:28]

Nevertheless, at least in the spoken language, we also find numerous examples in which this holds for animates as well. In some cases, plurality is indicated by some other means, such as a quantifier:

22. *ro sou³b bhai=ga ho=ghay=ga enem qa[ʔ]*
 and all brother=FOC that=way=FOC without water
uq=ga eṅ del=ki=may.
 drink=FOC return come=MID.PST=3PL
 'And *all brothers* thus returned, without drinking water.' [AK, 1:51]

In other cases only context can determine the correct interpretation, especially when the TAM/PERSON-syntagma is also not explicitly marked for plurality:⁴

23. *laʔ anij=te anij=aʔ purkha olsoʔol=oʔ. laʔ*
 then 1PL.INCL=OBL 1PL.INCL=GEN ancestor bring=ACT.PST then
ho=ki=yaʔ thoṅ=jo yad karay=na anij=aʔ
 that=PL=GEN for=ADD memory do=INF 1PL.INCL=GEN
bahut jaruri heke.
 great necessity QUAL.PRS
 'Then *our ancestors* brought us [here]. Thus, it is very important that we remember them.' [MS, 1:159]

We also find cases in which a Case-syntagma is marked for plurality, whereas the TAM/PERSON-syntagma is not:

24. *ho=ki bohut dinu aw=ki ro moṅ=ga*
 that=PL many day stay=MID.PST and one=FOC
beʔt=qom aw=ki.
 son=3POSS QUAL=MID.PST
 '... *they stayed* many days and they had (= there was) one son.' [MS, 1:21]

⁴ It is interesting to note in this example that the proform *ho=ki=yaʔ*, which refers to the same group of people as *purkha* 'ancestor[s]', is marked for plurality. Apparently, proforms are more prone to overt number marking when they refer to a non-singular entity than contentive morphemes.

In sum, although Case-syntagmas which refer to non-singular inanimate entities do show a greater tendency to lack overt number marking than those referring to animate entities, both types of Case-syntagmas can lack overt number marking and, *pace* Biligiri, both types are also compatible with overt number marking, at least in our corpus.

Note that there is a classifier, *j(h)an*, which is restricted to human reference (5.9.2), which would seem to suggest that, while there may not be a [\pm animate] distinction in the lexicon, there is perhaps a [\pm human] distinction. This argument is further strengthened by the fact that there is a similar distinction in the “pronominal” system (5.6).

However, the evidence for such an analysis is not entirely convincing: First, the classifier *j(h)an* can always be replaced by the more common general classifier *=o*, which is compatible with human, non-human animate, and non-animate reference, and classifiers are never obligatory. In addition, the [\pm human] distinction in the “pronominal” system is not strictly observed: [+human] forms can be used in reference to non-human entities and [-human] forms are also occasionally found with reference to humans (see 5.6 for examples).

There is thus no clear-cut division of the lexicon into [\pm animate] or for natural gender which is relevant for any morphosyntactic processes. While there is some evidence for a [\pm human] distinction, this distinction is not consistently observed.

5.1.2 *Definiteness and indefiniteness as grammatical categories*

Kharia has no definite articles. To indicate that a referential expression is to be interpreted as definite there are a number of strategies:

- The use of the oblique case marker with definite direct objects, which is restricted to Case-syntagmas with definite reference. With “indirect” objects and adverbials, however, its use is more or less mandatory (5.2);
- The use of demonstratives (5.8);
- Word order, with topical, generally definite information preceding focused information (7.7).

On the other hand, to refer to an unspecified, non-definite referential entity, *mon* or *ek* ‘one’ is used similarly to an indefinite article in English, as in the following example, where the cave is being referred to for the first time.

25. *sou^ʔb se maha be^ʔt=ɖom simɾa, moŋ maɾa=te ɖa^ʔ kuy=o^ʔ.*
 all ABL big son=3POSS Simra one cave=OBL water find=ACT.PST
 ‘His eldest son, Simra, found water in *a* cave.’ [AK, 1:22]

In this example, the cave is being introduced into the narrative. Although it is not mentioned again for several lines of text, in line [AK, 1:34] it is mentioned again (as well as repeatedly after that), where it is not marked by *moŋ* or *ek* and is marked as definite through the use of the demonstrative *ho* ‘that.MEDIAL’:

26. *tay beghma col=ki ro ho maɾa bo^ʔ=te*
 then Beghma go=MID.PST and that cave place=OBL
ɖam=ki.
 arrive=MID.PST
 ‘Then . . . Beghma went and he arrived *at that* cave.’ [AK, 1:34]

5.2 CASE

Case is denoted in Kharia by enclitics which attach directly to the right-most element of the semantic head of a Case-syntagma. There are three morphological cases:

- Direct* (zero marking)—the case of subjects, non-definite countable objects, and non-countable objects;
Oblique (marked by =*te*)—marks definite objects, “indirect objects” and some adjunct types.
Genitive (marked by =(y)a^ʔ)⁵—the case in which a semantic base appears / can appear (see discussion below) when it functions as an attribute in a larger semantic base.

Note that a number of temporal and locative deictics have special genitive forms:

⁵ From a historical perspective, =a^ʔ is the original form of this suffix, which is still found in most north Munda languages. The form =ya^ʔ results from the insertion of a non-phonemic glide in certain phonological environments which has become more or less mandatory and spread to other environments (see Section 2.4). Other allomorphs are =wa^ʔ (after /o/ and /u/), =^ʔ after /a/, =na^ʔ occasionally with demonstratives and =ga^ʔ in a number of deictic expressions but, with the exception of these last two, in all these environments as well =ya^ʔ is the most common form.

27. <i>ida?</i> ‘yesterday’	<i>id=ga?</i> ‘yesterday’s’
<i>musa</i> ‘today’	<i>mus=ga?</i> ‘today’s’
<i>tuḍa</i> ‘tomorrow’	<i>tuḍ=ga?</i> ‘tomorrow’s’
<i>hote</i> ‘there’	<i>hote=ya?</i> , <i>hot=a?</i> , <i>hot=ga?</i>
<i>ute</i> ‘here’	<i>ute=ya?</i> , <i>ut=a?</i> , <i>ut=ga?</i>

Case markers otherwise show no irregularities and simply attach to the rightmost element of the semantic base, regardless of its status.

Direct vs. oblique case

Consider the following example. Here the subject is morphologically unmarked (= direct case) while the oblique marker is used to signal both the locative adjunct (*ho maṛa bo?*=*te*) as well as the definite direct object (*ḍa?*=*te*):

28. <i>la?</i>	<i>poṛha=jo</i>	<i>col=ki</i>	<i>ro</i>	<i>yo=te,</i>	<i>ho maṛa</i>
then	Porha=ADD	go=MID.PST	and	see=ACT.PRS	that cave
<i>bo?</i> = <i>te</i>	<i>ḍam=ki</i>	<i>ḍa?</i> = <i>te</i>	<i>yo=yo?</i> .		
place=OBL	arrive=MID.PST	water=OBL	see=ACT.PST		
‘Then <i>Porha</i> also went and looks, he arrived <i>at the cave</i> [and] saw <i>the water</i> .’					
					[AK, 1:39]

The use of the oblique marker is somewhat speaker-specific. This is especially true with respect to “mass nouns” or non-countable entities. For example, for many speakers, *ḍa?* in the last example would not be marked for the oblique marker =*te* as it is not a countable entity.

As noted, the oblique is also used to refer to what translate as indirect objects (7.1) as well as certain adjuncts. Generally, if the recipient or goal of a trivalent predicate is mentioned and marked for the oblique, the direct object is not marked for the oblique case, as in the following example:

29. <i>raksin</i>	<i>gam=te</i>	<i>“ol=e</i>	<i>u</i>	<i>beṭa</i>	<i>ip=te,</i>	<i>juḡhay=ga</i>
witch	say=ACT.PRS	bring=ACT.IRR	this	boy	1sg=OBL	much=FOC
<i>sa?dhe=si?</i> .”						
torment=PERF						
‘The witch says “Bring <i>this boy to me</i> , he has tormented me so very much.”’						
						[BB, 1:81]

However, the recipient / goal as well as the direct object may both be marked for the oblique case:

30. *raʔa saŋgo=ɖom=te sob kayom=te utun=oʔ.*
 Rata friend=3POSS=OBL all matter=OBL relate=ACT.PST
 'Rata related *the entire matter to his friend.*' [BB, 2:55]

Also, although locative adjuncts are generally marked for the oblique case, as in (28) above, as the following shows, this is not obligatory. Also, temporals are rarely marked for the oblique case:

31. *ho=kaʔ din bheir u ɖair han ɖair eboʔ*
 that=SG.HUM day entire this branch that branch play
loʔ=na laʔ=ki.
 ITER=INF IPFV=MID.PST
 'All day he continually played *on this branch and that.*' [TK, 2:4]

Thus, the direct case is underspecified with respect to function, whereas the oblique case is positively marked in this respect.

In addition to the functions of the oblique case mentioned above, this case is often found with an instrumental interpretation in place of the more common postposition *tay* 'ABL; INST', which is specialized for this function and ablativity.

32. *iskul melay=kon arkhi golanʔ=te=ga leʔo aw=na*
 school stop=SEQ liquor rice.beer=OBL=FOC very.drunk QUAL=INF
laʔ=ki.
 IPFV=MID.PST
 'After quitting school he was usually heavily drunk *with liquor and rice beer.*' [RD, 2:32]

Finally, there are a number of examples in the corpus where the oblique case marks the subject of the clause. Some of these, such as the following example, can perhaps be considered production errors:

33. *raʔa=te kontheɖ bajhay=na=yaʔ mon um karay=na*
 Rata=OBL bird trap=INF=GEN mind NEG do=INF

laʔ=ki.

IPFV=MID.PST

‘*Rata* did not enjoy (= do the mind of) trapping birds.’ [BB, 2:17]

Here we can probably assume that the author originally intended to use the predicate *mon aw* ‘mind QUAL’, with approximately the same meaning as *mon karay* ‘enjoy’ but which requires an experiencer in the oblique case (6.3.4.2), but then used the predicate *mon karay*, which takes a direct-case subject.

There are, however, a number of cases in which this analysis is less certain, as in the following:

34. *mudui=ki=te=ko daw kuy=oʔ=ki.*

enemy=PL=OBL=CNTR chance find=ACT.PST=PL

‘*The enemy* found an opportunity.’ [MT, 1:93]

Although seldom, such examples do occur and, since they are generally restricted to the two southern-most speakers, it may well be that there is a certain regularity behind the use of the oblique case with subjects which is not yet clear and requires further study.⁶

Direct case vs. genitive

As noted above, the genitive is used to incorporate a semantic base as a modifier into another, larger semantic base. Consider the following example:

35. *ho raʔ-raʔ ghāṛi junay=kon lebu=ki koŋ=oʔ=may*

that blossom-RDP smell smell=SEQ person=PL know=ACT.PST=3PL

no u daru kuda koloŋ=yaʔ daru heke.

CMPL this tree millet bread=GEN tree PRS.COP

‘When they smelled the smell of the flowers, the people knew that

this tree is a *millet bread* tree.’ [BB, 1:20]

Here, *kuda koloŋ* is marked for the genitive as part of the larger semantic base. The fact that *raʔ-raʔ* is not marked for the genitive could perhaps

⁶ Note e.g. that the use of the cognate form *te* in Santali, which Neukom (2001) terms the instrumental / allative, is in fact productively used for “emphasis of the subject” (Neukom, 2001: 27f.).

be considered a kind of “haplogy” for the expected—and fully grammatical—form *raʔ-raʔ=yaʔ* ‘blossom-RDP=GEN’, although this is purely speculative.

Similar to the discussion above in which the direct case is found in a number of environments in which we would normally expect the oblique marker =*te*, the genitive is also often not found in environments where it might be expected, although the reasons for this are not entirely clear. Consider the following example:

36. *kongher merom gupa=na melay=kon, kuda koloŋ daru=te=ga*
 boy goat shepherd=INF leave=SEQ millet bread tree=OBL=FOC
iʔib tunboʔ aw=na laʔ=ki.
 night midday live=INF IPFV=MID.PST
 ‘The boy gave up shepherding the goats and lived day and night only
 in the *millet bread tree*.’ [BB, 1:25]

In (36) we find the exact same entity being referred to by the same author as in (35), but this time the semantic base is not marked for the genitive but rather appears unmarked for case, i.e., in the “direct case” In fact, it is quite common for semantic bases to appear without genitive marking when incorporated into a larger semantic base. Thus, here once again the direct case is not only morphologically unmarked, it is also unmarked with respect to function: While the function of the genitive is unambiguous when it is found, speakers obviously have a certain degree of freedom as to when to use it. When “possession” (in the sense of ownership) is intended, the genitive always appears to be used.⁷ However, when the relationship between the incorporated semantic base and the superordinate semantic base is not one of possession in the usual sense, the use of the genitive is apparently not obligatory.

As noted above, case markers attach directly to the last element of the semantic base. This differs from the adpositions (5.3), which either always take the genitive or which take the genitive with personal forms and proper names:

37. *oʔ=te* ‘the house (OBJ); in the house’ CASE MARKER
 house=OBL

⁷ Or one of the markers of inalienable possession, see Section 5.5.

38. *og=aʔ* bahare ‘outside the house’ POSTPOSITION
 house=GEN outside

There is only one construction in which the oblique marker =*te* can co-occur with the genitive: If the lexical head of a Case-syntagma phrase is not overtly expressed, as its identity is already known or is considered unimportant, the oblique marker =*te* attaches to the right-most element of the remaining semantic base, regardless of its status. If this element is a genitive determiner, this results in apparent “double case” marking. It is especially common with predicative alienable possession and with the meaning ‘at the home / place of’:

39. *ijɪ=aʔ=te* *saykal* *ayɪʔj*.
 1SG=GEN=OBL bicycle QUAL.PRS
 ‘I have a bicycle.’

This may be considered a “reduced” form of the following construction, in which the lexical head is overtly mentioned:

40. *ijɪ=aʔ* *boʔ=te* *saykal* *ayɪʔj*.
 1SG=GEN place=OBL bicycle QUAL.PRS
 ‘I have a bicycle.’ (literally: ‘There is a bicycle *at my place*.’)

The same is also true of the meaning ‘at the home / place of’:

41. *am=aʔ=te* = *am=aʔ* *boʔ=te* ‘at your place’
 2SG=GEN=OBL 2SG=GEN place=OBL

As the lexical head, *boʔ* ‘place’, is omitted in (39) and (41), the case marker =*te* simply attaches to the right-most element of the remaining semantic base, as was the case in (17) with respect to number marking.⁸

⁸ Malhotra (1982: 110) provides an example in which the oblique marker =*te* attaches to the genitive in a so-called “dative-subject” environment: *ijɪ=aʔ=te lemed laʔ=ta* [1SG=GEN=OBL sleep EMOT=MD.PRS] ‘I feel sleepy.’ Forms such as these are exceedingly rare and I have no such forms in my own corpus. Most important for the dialect I am describing here however is the fact that these examples were rejected by speakers I consulted. Hence I will consider them dialectal or perhaps idiolectal forms. Examples such as these may, however, be remnants of a postpositional origin of =*te*.

5.3 ADPOSITIONS

Adpositional phrases in Kharia consist of adpositions, generally postpositions, and a semantic base, often marked for the genitive, as in the following example:

42. *ro gam=o?* *no* "iⁿ=te *konon konon daru=ya?* *konon*
 and say=ACT.PST Cmpl 1sg=OBL small REP wood=GEN small
kundu?=ki=ya? *thoŋ khelawna bay=na sikhay[=e].*"
 child=PL=GEN PURP doll make=INF teach=ACT.IRR
 'And he said "Teach me how to make small wooden dolls *for young*
 (= *small*) *children.*"'

[BB, 2:56]

Kharia has a large number of postpositions and one preposition. There are also a number of elements which show many similarities to postpositions but which have only partially grammaticalized.

Adpositions in Kharia differ from case markers in that adpositions generally require the genitive case when the semantic head is a proform or proper name, otherwise the direct case. Case markers, on the other hand, attach directly to the last element of the semantic base and never take the genitive.⁹ In the following the most common adpositions are discussed as well as a number of forms which have not fully grammaticalized into adpositions.

enem 'without'

This is the only preposition in the language. Except in its depictive use (7.5.1.6) and when used to negate the masdar (see below) or a sequential converb (6.6.2.3), a semantic base always appears in the genitive, at least in our texts and in the first half of those in Pinnow (1965a):

43. *ani jahã ti²j col=ta=niŋ. u=te=ko*
 come.on INDEF side go=Mid.PRS=1PL.INCL this=OBL(='here')=CNTR

enem qa?=ya? go²j go²d=na=niŋ.
 without water=GEN die C:TEL=Mid.IRR=1PL.INCL

'Come on! Let's go somewhere. Here we will die *without water.*'

[TK, 1:7]

⁹ Unless of course this element itself is already marked for the genitive for other reasons. See 5.2, (39)–(41).

enem is also used to negate a secondary (or “non-finite”) predicate whose semantic head is a masdar (here: *aw-aw*).¹⁰ The entire Case-syntagma here appears in the oblique case to denote its adjunct status. The use of the oblique marker =*te* here is thus not connected to *enem* but to the adjunct status of this secondary predicate.

44. *enem* *raja* *ro* *rani=kiyar=a?* *aw-aw=te*
 without king and queen=DU=GEN QUAL-RDP=OBL
- kharīya=ki kaṭi?j=ko* *qher=ga akbakay=ki=may.*
 Kharia=PL somewhat=CNTR very=FOC be.in.a.flurry=MID.PST=3PL
 ‘Without there being a king and queen, the Kharia panicked somewhat.’
 [MT, 1:104]

The postpositions

There are simple and complex postpositions in Kharia. Most simple postpositions are not further analyzable from a synchronic perspective. A few of these, such as *lekhe* and *motabik* in the following list, have been borrowed from Indo-Aryan.

45. *buŋ* ‘with; INST; COMMITATIVE’ *motabik* ‘according to’
gha²d ‘for; PURP’ *tay* ‘ABL’
lekhe ‘like, similar to’ *seŋ* ‘before’ (TEMP)
lo²dho ‘after’ *thom, thoŋ* ‘for; PURP’

To this *jou* (seldom: *jaw*) ‘up to, as far as, until’ can tentatively be added. However, in all examples in our corpus, *jou* never takes the genitive case as do other postpositions, suggesting that it may in fact be a case marker, although this could also be due to the fact that it does not occur anywhere in our corpus with proforms. Its status as a postposition is thus somewhat questionable and requires further study.

As already noted, these postpositions generally take the genitive with proforms and proper names and the direct case with other semantic bases. This suggests that the relevant criterion is definiteness.

46. *ji=a?* *buŋ* ‘with me’ *ho=ka²=a?* *tay* ‘from him/her’
 1SG=GEN INST that=SG.HUM=GEN ABL

¹⁰ See 6.6.2.1. Note the use of the genitive here (*rani=kiyar=a?*) is not related to the preposition *enem* but marks the “subject” of the masdar.

kinir tay 'from a/the forest'
 forest ABL

However, this is not a hard-and-fast rule, as the following example shows, although it is still compatible with an analysis in terms of definiteness (due to the presence of marking for inalienable possession):

47. *janam=no=m=a?* *seṅ=ga* *oh re!* *ray=si?q=ijn*
 birth=2poss=2sg=GEN before=FOC INTERJ choose=PERF=1sg

am=te.

2sg=OBL

'Ah! *Even before your birth* I chose you.' [RK, 12:4]

Cf. also examples (223) and (224) on pages 204–205 for a contrast of the presence and absence of the genitive case with the same postposition (*lekhe*), corresponding to a definite/indefinite interpretation.

Complex postpositions generally derive from oblique-case relational semantic heads, whereby the oblique marker is generally no longer obligatory. Many of these forms derive from Indo-Aryan:

48. *badte* 'after' (IA) *mo?jhi(=te)* 'amidst, among' (IA)
bahar(=te) 'outside of' (IA) *mugam(=te)* 'in front of'
bhitare(=te) 'within' (IA) *toblun(=te)* 'on top of, above'
kundā²b(=te) 'behind' *tuta(=te)* 'under, below'

As with the simple postpositions, complex postpositions generally take the genitive with proforms and proper names, but show a higher tendency to take the genitive with other semantic bases as well.

49. *sadhu=ki=ya?* *mo?jhite* 'amongst the holy men'
 holy.man=PL=GEN amongst

They do not require the genitive, however: *tunbo? mo?jhite* 'in the middle of the day' Consider also the following example:

50. *tama u eriya=te uṛisa=te=jo samay goṛ=ki=may*
 now this area=OBL Orissa=OBL=ADD meet C.TEL=MID.PST=3PL
 u bhitarete *samay=ki=may.*
 this within meet=MID.PST=3PL

‘And now, in this area, and in Orissa as well, they met up with one another, here (= *within this* they met).’ [MS, 1:235]

There are also at least two postpositions which have evolved from contentive morphemes denoting ‘place’ and the simple postposition *tay* ‘ABL’. The ablative meaning seems to have been lost entirely here:

51. *bo?tay, kho?tay* ‘up to, until, as far as’ (cf. *bo?* ‘place’, *kho?* ‘place’)

52. *ho=kaṛ* *ḍisa?* *ḍisa?* *kho?tay=jo* *merom* *gupa=na*
that=SG.HUM far REP up.to=ADD goat guard=INF

la?=kho?

IPFV=PST.II

‘He used to tend the goats, even going (= *also up to*) very far.’

[RD, 1:4]

Alternatively, the form *kho?te*, with the same meaning, is also found:

53. *ina no le?bḍom, gupa kaṛ, diyo=ga aḍi=ya?* *merom=ki=te*
because boss guard person daily=FOC ANAPH=GEN goat=PL=OBL

katib=na=? *thon* *pōga* *bajay=na* *la?=kho?*
gather-<CAUS>=INF=GEN PURP horn sound=INF IPFV=PST.II

ro soub bheṛi merom=ki le?bḍom=ya? *kho?te* *ḍel* *kan=na*
and all sheep goat=PL boss=GEN up.to come CONT=INF

la?=kho?=may.

IPFV=PST.II=3PL

‘Because the boss, the shepherd, would daily sound the horn to gather his goats, and all the sheep and goats would come *to* the boss.’

[RD, 1:12]

54. *ab=a?* *raksin aḍi=ya?* *gone ro ramaḍ=ki=te cokhay=o?*
NOW=FOC witch ANAPH=GEN tooth and nail=PL=OBL sharpen=ACT.PST

ro daru ḍeb=na thon sumbo? kho?=te col=ki.

and tree ascend=INF PURP base place=OBL go=MID.PST

‘Now the witch sharpened her teeth and (finger)nails and went *to the base* [of the tree] to climb it.’

[BB, 1:36]

Cases such as these last two examples are perhaps better considered oblique-case marked Case-syntagma, where the lexical head of the semantic base is the lexical morpheme *kho?* 'place', as suggested by the glosses.¹¹

There are also a number of elements which show some similarities to postpositions while retaining to some degree the characteristics of the categories from which they have developed, whether a non-finite TAM/PERSON-syntagma or a Case-syntagma in adjunct function:

"Postpositions" from converbs

Kharia has (at least) the following three forms which, while retaining their original meanings in most contexts, have also taken on a new, grammaticalized meaning. Incidentally, all three examples have corresponding forms in a large number of other South Asian languages.

hoy=kon 'via, by way of'

55. *ho=ki iran balucistan hoy=kon dhirom dhirom*
 that=PL Iran Baluchistan become=SEQ slowly REP

dam=ki=may.

arrive=MID.PST=3PL

'They arrived slowly *via* (= *having become [in]*) Iran and Baluchistan.'
 [MT, 1:36]

melay=kon 'other than, except for'

Said by a dog in a children's story who had been looking for a friend among the animals of the forest but then realized that no animal was as worthy of being his friend as people. Note that *lebu=te* is marked for the oblique case, showing that this is not a true postposition.

56. *u dumiyã=te lebu=te melay=kon iŋ=a? sango bone=na*
 this world=OBL man=OBL leave=SEQ 1SG=GEN friend become=INF

laik ber=jo hoqom umborij=may.

worthy who=ADD other NEG.QUAL.PRS=3PL

¹¹ See below, "Postpositions from adjuncts". However, unlike *bo?* 'place', which is discussed in that section, *kho?* 'place' does not occur in my data in conjunction with the plural marker and seems to have a different status than *bo?*. This issue requires further study.

‘In this world *other than* (= *having left*) people there are no others worthy of becoming my friend.’ [BB, 3:67]

dhoʔ=kon ‘with’

In addition to its original meaning ‘having taken’, *dhoʔ=kon* (or here the dialectal variant *dhoʔ=kan*) can have an instrumental or comitative meaning. Note that here, as with *melay=kon* above, the supposed “object of the postposition” is marked for the oblique case.

57. *ro ho naw beʔa kimin=ki=te=ga* *dhoʔ=kan*
 and that nine son daughter.in.law=PL=OBL=FOC *grab=SEQ*
hindustan... ʔel=ki=ki.
 India come=MID.PST=PL
 ‘and *with* (= *having taken*) those 9 sons [and 9] daughters-in-law, they came to India.’ [MS, 1:58]

“Postpositions” from adjuncts

There are two locative markers, *boʔte* ‘LOC’ and *hinte* ‘LOC’, and one allative marker, *tiʔj* ‘side, direction’ which, in addition to their original meaning, also have many of the characteristics of postpositions.

boʔte

boʔ(te), literally ‘at the place (of)’, commonly functions as a general locative marker, as in the following example. In examples such as these, it is debatable as to how much of the original lexical meaning of this morpheme remains.

58. *jahāy kon=ga ho daru boʔ=te*
 INDEF.HUM INDEF=FOC that tree place=OBL
ʔam=na laʔ=ki=may.
 arrive=INF IPFV=MID.PST=3PL
 ‘Some people began to come to *that tree*.’ [BB, 1:26]

As can be seen in the last example, *boʔ* is glossed here as ‘place’ This is primarily because it is often difficult to determine the extent to which it has lost its lexical meaning. For example, (58) could easily be analyzed as having a locative adjunct, *boʔte*, the (adverbial) Case-syntagma, which is modified by *ho daru* ‘that tree’, i.e., ‘that tree place’ It could, however, also be analyzed as a postposition.

However, this latter analysis does not always seem possible, as in the following examples, due to the presence of plural marking on *bo?*:

59. *suru=te=ko kayam=na umay lam=te*
 beginning=OBL=CNTR speak=INF NEG.3PL seek=ACT.PRS
lekin ab ina no iskul bo?=ki=te=jo ab kharīya=ya?
 but now because school place=PL=OBL=ADD now Kharia=GEN
paṛhai suru hoy=taʔj.
 teaching beginning become=MID.PROG
 'At first, they don't want to speak [Kharia] but now because, at schools, too, the teaching of Kharia is beginning.' [AK, 5:23]

60. *khōṭi poʔda bo?=ki=te tam jou kharīya=ki*
 village.section village place=PL=OBL now up.to Kharia=PL
kaḍom cakḥna[ʔ] hinte saṅsaṅ umay may=te.
 fish curry LOC turmeric NEG.3PL MIX=ACT.PRS
 'Up to the present day, in the villages and village sections, the Kharia do not mix turmeric into fish curries.' [MT, 1:115]

There are often almost identical contexts in the corpus in which the same entity is referred to once with and once without *bo?*, strongly suggesting that the use of *bo?* is entirely facultative:

61. *jhaṛi nāw=o kulam hoy=ke apan apan gotor*
 all nine=CLASS brother become=SEQ REFL REP clan
no?=na parej guʔd=ga u thāṛo bo?=ki=te
 eat=INF abstain like=FOC this place place=PL=OBL
ḍoko=ki=may.
 sit.down=MID.PST=3PL
 'They all became nine families (= brothers), each abstaining from eating [the animal of] its own family [i.e., their totem animal, and] settled in these places.' [MT, 1:247]
62. *u thāṛo=ki=te=jo purkha=ki umay sekḥ=oʔ.*
 this place=PL=OBL=ADD ancestor=PL NEG.3PL be.held=ACT.PST
 'In these places as well the ancestors could not be contained [i.e., they were too numerous].' [MT, 1:248]

As (59), (60) and (61) show, *bo?* can also retain all of the morphological marking of a semantic base. This strongly suggests that it is not a postposition, despite the “semantic bleaching” which has taken place.

hinte

Similarly, *hinte*, which derives from *hin* ‘that (DISTAL)’ and the oblique case marker =*te*, originally with the meaning ‘there, at that [place]’, has developed in the speech of many speakers to a (pseudo-)locative postposition. Unlike *bo?te* above, however, in which *bo?* ‘place’ originally functioned as the semantic head of a Case-syntagma, *hinte* would appear to have originally been used in appositive function with the meaning ‘there’ (cf. also *hinte* in (60) above).

63. *modi=ya?* *konon* *beṭ=ḍom=te* *ḍain=ki* *kaṭa* *hinte*
 Modi=GEN small son=3POSS=OBL witch=PL foot LOC

ban tar=o?=ki.

spell beat=ACT.PST=PL

‘Witches cast (= beat) a spell on Modi’s youngest son’s feet.’

[RD, 2:103]

64. *ki?jte* *no ki?jte* *memon* *jou* *khariya* *maha* *maha*
 how.much or how.much year up.to Kharia big REP

rokedḍ hinte sanḡodḍ=ki=may.

sand LOC walk=MID.PST=3PL

‘For a couple of years the Kharia wandered in the great desert.’

[MT, 1:6]

65. *raṭa jal=te theila hinte sajay=kon kundab=te*
 Rata net=OBL (hand)bag LOC put.into=SEQ back=OBL

laḍdhe=kon kinir col=ki.

load=SEQ forest go=MID.PST

‘Rata put the net into the handbag, loaded it on his back and went to the forest.’

[BB, 2:24]

Note that *hinte* in examples such as these is glossed as ‘LOC’ and not as ‘there’ This is due to the fact that, unlike *bo?te*, the semantic contribution of *hinte* is always minimal in such examples. In addition, were it not present, there would be a locative adjunct unmarked for case, although locatives are otherwise generally marked for their locative function. Thus *hinte* would seem to have grammaticalized to a greater extent than *bo?te*,

although it is also found in its original function as a deictic locative with the meaning ‘there, at that [place]’

tiʔj

tiʔj ‘side; direction’ has grammaticalized to some extent so that it now often functions as a kind of allative postposition in many contexts:

66. *gam=oʔ ro raksin jhēriya tij col=ki.*
 say=ACT.PST and witch small.river side go=MID.PST
 ‘The witch said [this] and went *towards* the river [to wash for dinner].’
 [BB, 1:66]

67. *modi gitaʔ idjib=te peʔ och=oʔ ro moʃn*
 Modi lie.down night=OBL cooked.rice take.out=ACT.PST and one
loʔa ɖaʔ aʔɖ=kiyar=yaʔ gitaʔ koʔhri
 cup water ANAPH=DU=GEN lie.down room
tiʔj ɖi<ʔbh>ar=oʔ.
 side enter-<CAUS>=ACT.PST
 ‘At bed-time Modi took out some cooked rice and a cup of water and brought them *to* the couple’s (= their (DU)) bedroom.’
 [RD, 2:59]

As the examples show, *tiʔj* is glossed as ‘side’ This is a somewhat arbitrary decision, as it is not clear to what extent it has lost its lexical content in these examples. At any rate, as the following examples show, it clearly does not form a compound with the lexical head of the semantic base of the Case-syntagma, as the semantic head in the following two examples is separated from *tiʔj* by plural marking:

68. *kinir jhanʔkor biru=ki tiʔj paro=ga ho=kaʔ rel*
 forest ECHO mountain=PL side cross=FOC that=SG.HUM train
sarʔak tiʔj col=ki.
 road side go=MID.PST
 ‘[When] crossing *towards* the forests and mountains, he followed (= went [on] the side [of]) the train tracks.’
 [RD, 1:6]

69. *u boʔ=ki tiʔj=ga ciʔna=ki buli khor=na laʔ=ki=may.*
 this place=PL side=FOC chick=PL wander ITER=INF IPFV=MID.PST=3PL
 ‘The chicks were all running all over the place here (= *towards* these places).’

Finally, note that *tiʔj* very occasionally does not refer to movement towards a goal but appears to have a purely stative locative interpretation:

70. *gupa dāṭ tij kongher moṇ gaḥṭa lay=sikh=oʔ,*
 shephard field side boy one hole dig=PERF=ACT.PST
ho=te=ga koloṇ=te til=na laʔ=ki.
 that=OBL(= 'there')=FOC bread=OBL bury=INF IPFV=MID.PST
 'The boy had dug a hole *at* the grazing field, on that very spot he
 used to bury the bread.' [BB, 1:15]

5.4 NUMBER

Kharia has three grammatical numbers, both on Case-syntagmas and TAM/PERSON-syntagmas: Singular (unmarked), dual (marked by =*kiyar*) and plural (marked by =*ki*).¹² Number marking for non-singular entities is optional in the spoken language, especially if these are non-human. In written texts, number marking is generally considered obligatory, even if duality or plurality is indicated elsewhere in the clause.

71. *lebu (= lebu=∅)* 'man, person' *lebu=kiyar* 'two men, two people'
lebu=ki 'men, people'

Number is often marked only once per clause in the spoken language—either on a Case-syntagma or on the TAM/PERSON-syntagma (for subjects), but often not on both:

72. *ho=ki boḥut dinu aw=ki ro moṇ=ga beʔt=ḍom*
 that=PL many day stay=MID.PST and one=FOC son=3POSS
aw=ki.
 QUAL=MID.PST
they stayed many days and they had one son.' [MS, 1:21]

73. *u=ghay andaj ḍom=ta no ata bhere israeli lebu=ki*
 this=way guess PASS=MID.PRS CMPL Q time Israeli person=PL

¹² On the relationship between =*ki* and =*kiyar* (< **ki=bar* 'PL=two'?) see Pinnow (1966: 164).

rusuŋ samudar=te paro=na laʔ=ki, *hin bhere=jo*
 red ocean=OBL CROSS=INF IPFV=MID.PST then time=ADD

khariya=ki ho boʔ=ki=te=ga aw=ki=may.
 Kharia=PL that place=PL=OBL=FOC QUAL=MID.PST=3PL

‘Thus it is assumed (= guessed) that at the time that *the Israelis were crossing* the Red Sea, at that time the Kharia were at those places as well.’ [MT, 1:22]

Compare also the occurrence of *khariya=ki* in example (73) with the form *khariya* in the next example. Here, only the TAM/PERSON-syntagma indicates the plurality of the subject.

74. *laʔ ho=te khali khariya bheir aw=ki=may*
 then that=OBL(= ‘there’) only Kharia much live=MID.PST=3PL

oqoʔ hoqom jait=ko ... ho=ki kuday=ga
 and other ethnic.group=CNTR that=PL chase=FOC

ol=teʔj=ki

V2:bring=ACT.PROG=PL

‘Then only *the Kharia lived* there and they drive away the other ethnic groups.’ [MS, 1:153]

The “singular” is thus both morphologically and functionally unmarked. A Case-syntagma unmarked for number may thus refer either to a singular or non-singular entity, whereas dual and plural marking are both positively marked.

“Dual” and “Plural”

In addition to their primary function of denoting duality and plurality, the markers referred to here as “dual” and “plural” also have a number of other functions, which we now turn to.

“Dual”

The “dual” in combination with a personal name refers to that person and a person who is related to him or her in some fashion, usually a husband or wife, as in the next examples.

75. *aʔb etwa=kiyar modi=yaʔ biha=yaʔ tiha=na*
 now Etwa=DU Modi=GEN marry=GEN arrange=INF

absiph=o?=kiyar.

begin=ACT.PST=DU

‘Now *Etwa and his wife* began to arrange for Modi’s wedding.’

[RD, 2:44]

76. *modi=kiyar=ya? lachan yo=kon aʔb ayo aba=dom=kiyar*

Modi=DU=GEN behaviour see=SEQ now mother father=3POSS=DU

ro kulam=dom=ki bikhre goʔq=ki=may.

and sibling=3POSS=PL annoy C.TEL=MID.PST=3PL

‘Seeing the behaviour of *Modi and his wife*, his mother and father and his brothers became annoyed.’

[RD, 2:88]

Another, more common use of the “dual” is to denote politeness or “honorificity” (HON). In the first and second persons, this use is restricted to singular entities, where the exclusive form of the first person, dual (*iyjar*) is used in reference to oneself when speaking to someone to whom respect is to be shown and the second person, dual (*ambar*), to address someone to whom respect is due. In the third person, on the other hand, the use of the dual marker is found in reference to both singular and non-singular entities. Both the TAM/PERSON-syntagma and a Case-syntagma are generally marked for the “dual”/honorific.

77. *ho=kaʔ=a? tay konon bahin=kiyar tama*

that=SG.HUM=GEN ABL small sister=HON now

intermiqiyat=a? paricha likha=te=kiyar.

intermediate=GEN exam write=ACT.PRS=3.HON

‘Her younger *sister* (HON) is now *writing* the intermediate exams.’

[AK, 4:12]

78. *ghaʔ aʔi je bhere eʔ=ki r[o] aba=dom=te*

therefore ANAPH CREL time return=MID.PST and father=3POSS=OBL

souʔb bhai=kiyar=a? sori? potom=te kholay

all brother=HON=GEN together bundle=OBL open

kholay ob-yo=na laʔ=ki=may, se bhere aʔi=ya?

REP CAUS-see=INF IPFV=MID.PST=3PL that time ANAPH=GEN

potom=te soreʔ kui=ki.

bundle=OBL stone find=MID.PST

'Therefore, when he returned, and along with all his [8] *brothers* (HON) they opened the bundles and were showing them to their father, at that time he [unexpectedly] found a stone in his own bundle.'

[AK, 1:73]

"Plural"

In addition to denoting more than two of a particular entity, the plural is also often used to denote approximation:

79. *u boʔ=ki tiʔj=ga ciŋna=ki buli khor=na laʔ=ki=may.*
 this place=PL side=FOC chick=PL wander ITER=INF IPFV=MID.PST=3PL
 'The chicks were all running all over the place here (= to these *places*).'

80. *ho=ki hephaʔd=ki=te=ga doko=ki=may.*
 that=PL vicinity=PL=OBL=FOC sit.down=MID.PST=3PL
 'They settled *in the (general) vicinity*.'

Other common examples: *idaʔ=ki* 'yesterday or so', *tuɖa=ki* 'tomorrow or so':

[MT, 1:246]

81. *ho=kaŋ idaʔ=ki del=ki hoy.*
 that=SG.HUM yesterday=PL come=MID.PST INFER
 'S/he came *yesterday (or so)*, I guess.'

Another common, related use is to denote "etc.", "or whatever":

82. *de babu, amoʔd gujuŋ=na roʔ ɖaʔ=ki uʔd*
 well boy wash.face wash.feet=MID.IRR FOC water=PL drink
ɖoŋ=e=m tay dukham sukham=na=pe.
 A:TEL=ACT.IRR=2SG.then chat=MID.IRR=2PL
 'Well then, boy, wash your face and feet. Then you will drink *water (etc., or whatever)* and then you will all have a chat.'

[Kerkettā, 1990: 26]

The plural can also be used to designate someone and his or her family members as a kind of collective:

83. *gomke=ki=yaʔ thoŋ konkeʔ romkub=yaʔ peʔ*
 master=PL=GEN for slender husked.rice=GEN cooked.rice

dui=yo sājḥ isin=na laʔ=ki.

two=CLASS evening cook=INF IPFV=MID.PST

‘For *the master [and his wife and brother]*, [a servant] cooked fine (= slender) rice at two o’clock in the afternoon (= evening).’

[BB, 1:9]

The plural is also occasionally found when referring to a (single) group of people (84) or to a group of objects (85) collectively:

84. *mudui=ki* ‘the enemy’

enemy=PL

85. *col=kon kaʔ kom, tarvair, mirjay ro dhoti=ki*
go=SEQ bow arrow sword weapon and dhoti=PL

ol=kon raja=te ter=te.

bring=SEQ king=OBL give=ACT.PRS

‘[He] goes, brings *the bow and arrows, sword, mirjai and dhoti* and gives them to the king.’

[Kerkettā, 1990: 5]

Finally, another means of denoting the non-singularity of an entity is through the repetition of a determiner or modifier, as in the following two examples.

86. *kinir=te kisim kisim=yaʔ konthed ro hoḍom jiwjantu=ki*
forest=OBL type REP=GEN bird and other animal=PL

aw=na laʔ=ki=may.

live=INF IPFV=MID.PST=3PL

‘In the forest lived *many different types* of birds and other animals.’

[BB, 2:5]

87. *aḍi=yaʔ rup buḥ=ga konon konon murti*
ANAPH=GEN form INST=FOC small REP statue

bay=oʔ.

make=ACT.PST

He made *small* statues in (= with) his [own] form.’ [AK, 3:10]

5.5 INALIENABLE POSSESSION

Kharia distinguishes morphosyntactically between alienable and inalienable attributive possession. With alienable possession, the possessor

Table 5.3: Inalienable possessive markers

	<i>Singular</i>	<i>Dual</i>		<i>Plural</i>	
1st person	(=na)= <i>ɟ</i> , (=na)= <i>ɟ̃</i> , (=na)=(i) <i>ɟ̃</i>	<i>Incl.</i>	<i>Excl.</i>	<i>Incl.</i>	<i>Excl.</i>
2nd person	= <i>nom</i>	(=na)= <i>naɟ</i>	(=na)= <i>jar</i>	(=na)= <i>niɟ</i>	(=na)= <i>le</i>
3rd person	= <i>ɟom</i>	(=no)= <i>bar</i>	= <i>ɟom</i> = <i>kiyar</i>	(=no)= <i>pe</i>	= <i>ɟom</i> = <i>ki</i>

appears in the genitive before the lexical head of the semantic base. With inalienable possession, instead of (or in addition to) a genitive attribute the lexical head itself is marked for the possessive relationship by one of the suffixes in Table 5.3, which refer to the possessor (adapted from Malhotra, 1982: 96).¹³

Malhotra (1982: 97) notes that the forms for the 3rd persons dual and plural are ambiguous—the number marking here may refer either to the possessor or to the possessed entity. Only context allows for the correct interpretation: *kulam*=*ɟom*=*ki* ‘their (PL) brother(s) / his/her brothers’

There are considerable differences of opinion with respect to these markers: Pinnow (1966: 159) lists merely one form each for the first, second and third persons, (=na)=*ɟ*, =*nom* and =*ɟom*, respectively, regardless of the number of the possessor. Similar data are found in Mahapatra (1976: 809). This was also confirmed by various speakers. Nevertheless, counter-examples are easy to find. Obviously, both systems are in use.

88. *gandur*=*na*=*niɟ* *laɟe*=*ki*, *laɟe* *dakha*=*ki*
Gandur=1POSS=1PL.INCL *fight*=MD.PST *fight* *try*=MD.PST

hãʔniɟ.

indeed

‘*Our Gandur* fought, he tried to fight, indeed.’ [Kerkeṭṭā, 1990: 8]

The inalienable construction may also be “reinforced” by the alienable construction. At present it is not clear whether this “reinforcement” fulfills any semantic or pragmatic functions: Although the genitive attribute can disambiguate otherwise ambiguous cases involving a third-person possessor, as in (89), a genitive attribute is also compatible with a lexical head marked for a second-person inalienable possessor, as in (90):

¹³ Note the similarity between many of these forms and the free-standing proforms, discussed in Section 5.6.

89. *ho=te=ga* *manu* *purkha* *buṛḥa=?*
 that=OBL(= 'there')=FOC Manu ancestor old.man=GEN

saw-ṛay=ḍom *jiyom* *ter=o?*
 spouse-woman=3POSS life give=ACT.PST
 'Right there, *the elderly ancestor Manu's wife* died.' [MT, 1:110]
90. *am=a?* *ḍhanḡar=nom* *gil* *ḍom=sikh=o?*
 2SG=GEN servant=2POSS beat PASS=PERF=ACT.PST
 'Your servant had been beaten.' [adapted from Abbi, 1993: 551]

As Malhotra notes, markers for inalienable possession are most commonly found with kinship terms, body parts and domestic animals. The following presents a few examples from our corpus:

91. *aba=ṇ* 'my father' 93. *ma=ḍom* 'his/her mother'
 father=1SG mother=3POSS
92. *haṛ=no=m* 'your bone(s)' 94. *bokob=ḍom* 'his/her head'
 bone=2POSS=2SG head=3POSS

In fact, of the 203 occurrences of the inalienable possessive marker in our own corpus, 165 (ca. 81%) refer to relatives such as *aba* 'father', *ma* 'mother', *beṭa* / *be't* 'son', *bi'j* 'daughter', etc. Other contentive morphemes referring to relatives can also be marked for inalienable possession. This contentive morpheme can also be ellipsed, as in the following examples, when context permits:

95. *maha=ḍom* 'the elder (son)', *nawa=ḍom* 'the ninth (son)'¹⁴
 big=3POSS nine=3POSS

In addition to kinship terms, there are also 7 cases in the corpus in which the possessive markers refer to body parts, such as *kaṭa* 'foot' or *pat(t)a* 'tail':

96. *ata* *raji* *co=na=m*, *sango*, *kaṭa=no=m=a?* *juta*
 Q country go=MID.IRR=2SG friend foot=2POSS=2SG=GEN shoe

¹⁴ There are also two unclear cases in [MS, 1: 197], where both *maha* 'big' and *konon* 'small' are marked for inalienable possession but do not appear to refer to relatives, although this is uncertain.

paŋj=te jumi likha ʃu=ye=m.
 footstep=OBL name write DPT=ACT.IRR=2SG

'To what country will you go, friend, you will write your name with your footprints (= the shoe footsteps *of your feet*) as you depart.'

[BB, 5:1]

97. *solo? lebu=ya? hepađ patta=đom=te hilay=ga col=ki*
 dog man=GEN vicinity tail=3POSS=OBL shake=FOC go=MID.PST

ro đoko gođ=ki.
 and sit.down C.TEL=MID.PST

'The dog went near the man, wagging *his tail*, and sat down.'

[BB, 3:59]

However, even with these content heads, inalienable possessive markers are by no means obligatory. While terms of kinship are almost always marked for inalienable possession, body parts generally are not:

98 *ba? ruŋku⁷b kay=na kayna ti? kaʃa loyo?=ki.*
 "rice" lift=INF REP hand foot tired=MID.PST

'She became tired of picking up the rice (= [*her*] *hands and feet* become tired, lifting the rice).'

[MT, 1:41]

There are also two almost identical examples in the corpus involving livestock marked for inalienable possession, as in the following example:

99. *bheŋi merom=đom=ki*
 sheep goat=POSS=PL
 'his sheep and goats'

In addition to kinship terms, body parts and livestock, there are 27 additional occurrences of inalienable possession in the data,¹⁵ most of which are compatible with the usual notion of semantic inalienability, such as one's village, one's master (of a servant), the meat (of an animal), birth, life and gods:

¹⁵ This is the total number of occurrences—many of the individual contentive morphemes occur more than once.

100. *poʔda=ɖom=ki* ‘their villages’
village=3POSS=PL
101. *gomke=ɖom=ki* ‘their masters’
master=3POSS=PL
102. *komanɯ=ɖom* ‘its (= the dead animal’s) meat’
meat=3POSS
103. *janam=no=m* ‘your birth’
birth=2POSS=2SG
104. *jiyom=na=iŋ* ‘my life’
life=1POSS=1SG
105. *bhagwan=ɖom=ki=te* ‘their gods (OBJ)’
god=3POSS=PL=OBL

Slightly less “prototypical” cases include terms for ‘friend’:

106. *sango=ɖom* ‘his friend’
friend=3POSS
107. *dular=na=iŋ* ‘my “loved one”’, i.e., ‘my friend’¹⁶
love=1POSS=1SG

There is also one case in which an article which is worn around the neck and over one shoulder, the *janew* ‘holy cord’, is marked for inalienable possession, although articles of clothing are not otherwise marked for inalienability in the texts.

108. *haɖa=na=?* *seŋ* *ho=kaɽ* *aɖi=ya?*
urinate=INF=GEN first that=SG.HUM ANAPH=GEN
- janew=ɖom=te* *ho* *daru=te* *taŋay*
holy.cord=3POSS=OBL that tree=OBL hang

¹⁶ *dular* refers to platonic ‘love’ or ‘friendship’ or ‘affection between relatives’, not ‘erotic love’

goʰh=oʔ *ro* *haqa=na* *col=ki*.

C:TEL=ACT.PST and urinate=INF go=MID.PST

'Before peeing, he . . . hung *his holy thread* . . . on that tree and went to pee.'
[AT, 2:11]

This can perhaps be explained as the *janew* is an important religious article worn by men of the priestly class and is an essential element of a priest's outfitting.

Other cases are found which may also be considered semantically inalienable as they necessarily involve a "possessor" of some kind, while not fitting into any of the categories mentioned above. For example, in (109) we find *najeir* 'glance, look (n.)' marked for inalienability, which can perhaps be explained since a glance cannot exist without its "possessor" or, rather, experiencer.

109. *raksin* *han=tiʔj* *u=tiʔj* *tuta* *tobluj* *najeir=ɬom=te*
witch that=side this=side bottom top glance=3POSS=OBL

ghumray=oʔ.

tum.around=ACT.PST

'The witch cast *her glance* this way and that, up and down.'

[BB, 1:33]

In a similar vein, *longoy=ɬom=ki* in (110) can perhaps be explained as a shadow necessarily requires something which produces the shadow, while *pap=le=te* in (111), taken from Pinnow (1965a), necessarily requires someone who commits sins:

110. *bonor* *daru* *bagoyca*, *bhore* *khirom* *lenge=taʔj*,
green tree garden full river flow=MID.PROG

longoy=ɬom=ki *sundar-boʔ* *luhur* *luhur* *koyo=te*.

shade=3POSS=PL beautiful-INTENS COOL REP wind=OBL

'The luxuriant green trees [of the] garden, the full river is flowing, *their shade* [is] very beautiful in the cool breeze.'
[AT, 2:2]

111. *pap=le=te* *chemay=e* 'forgive *our sin[s]*' [HJPa:256,1]
sin=1PL.EXCL=OBL forgive=ACT.IRR

Nevertheless, there remain a few cases which must be seen as “language-specific”, such as *bacan* ‘rain’ in the following example, from a traditional song. Here, =*nom* ‘your’ refers to God:

112. *o<ʔb>ɔɔr ɔɔr=e bacan=no=m=te.*
 hear-<CAUS> A: TEL=ACT.IRR rain=2POSS=2SG=OBL
 ‘Let *your* rain resound.’ [AT, 1:3]

Finally, it should be noted that no contentive morpheme in Kharia requires the presence of an inalienable possessive marker. Even contentive morphemes denoting ‘mother’ and ‘father’, which virtually always appear marked for inalienable possession, can appear without these markers.

113. *ro mane=te=may no u beɽa ro beɽi=kiyar=ga*
 and believe=ACT.PRS=3PL Cmpl this boy and girl=DU=FOC
loʔɔɔho mugam col=ke soub duniya=yaʔ lebu=ki=yaʔ
 after forward go=SEQ all world=GEN person=PL=GEN
ayo ro aba heke=kiyar.
 mother and father QUAL.PRS=DU
 ‘And they believe that this boy and girl later, having gone forth, are *the mother and father* of the people of the whole world.’ [AK, 3:18]

5.6 PROFORMS

The present section deals with proforms which correspond in function largely to pronouns in other languages.

Free-standing proforms

The free-standing proforms are presented in Table 5.4. The use of these proforms in Kharia is “optional” when the identity of the referent is not focused, i.e., Kharia is a “pro-drop” language.¹⁷

¹⁷ But see also the discussion of “arguments” in 7.1.1.

Table 5.4: The free-standing proforms

	<i>Singular</i>	<i>Dual / Honorific</i>		<i>Plural</i>	
		<i>Inclusive</i>	<i>Exclusive</i>	<i>Inclusive</i>	<i>Exclusive</i>
1	<i>ij̄n / iñ</i>	<i>anañ</i>	<i>ij̄ñjar</i>	<i>aniñ</i>	<i>ele</i>
2	<i>am</i>	<i>am=bar</i>		<i>am=pe</i>	
3	<i>ađi</i>	<i>ađ(i)=kiyar</i>		<i>ađ(i)=ki</i>	
	<i>ho=kaṭ, ho=je?</i>	<i>ho=kiyar</i>		<i>ho=ki</i>	
	<i>u=kaṭ, u=je?</i>	<i>u=kiyar</i>		<i>u=ki</i>	
	<i>hin=kaṭ, hin=je?</i>	<i>hin=kiyar</i>		<i>hin=ki</i>	
	<i>han=kaṭ,</i>	<i>han=kiyar</i>		<i>han=ki</i>	
	<i>han=je?</i>				

As noted in Section 5.4, the dual is also used to denote the honorific status of singular entities in the first and second persons. When referring to oneself to someone to whom respect is due, the exclusive form of the first person, *ij̄ñjar*, is used, whereas the second person, dual, *ambar*, is used for addressing a person to whom respect is due (for an example of this use with the second person, see *ambar* (115) in below). In the third persons, the dual may be used as an honorific marker for any number of entities:

114. *ij̄n=a?* *tay* *konon* *tin* *bhaya=ñ=kiyar*
 1SG=GEN ABL small [. .] three brother=1POSS=HON

ayi?j=kiyar.

QUAL.PRS=HON

'I have 3 younger brothers (= *my three . . . smaller brothers are*).'

[AK, 4:5]

Inclusive / exclusive

In the non-singular first persons, there is an inclusive/exclusive distinction. The inclusive denotes that the speaker includes the addressee in the reference ('we including you'), whereas with the exclusive, the addressee is excluded ('we without you').

In the following example from a children's story, a dog who has no friends has just approached a hare to ask him if he will be his friend:

115. *solo?* *gam=te* "*am=bar* *bura* *um=bar* *mane=te?j*
 dog say=ACT.PRS 2=2HON bad NEG=2HON consider=ACT.PROG

la? *anañ* *boriya* *sañgo* *bone=kon* *sori=ga*
 then 1DU.INCL both friend become=SEQ together=FOC

aw=na=naŋ.”

live=MID.IRR=1DU.INCL

‘The dog says “If you (HON) don’t consider it bad then *we two* (= you and I) can become friends and live together.”’ [BB, 3:10]

In contrast, in the following example part of the group—the speakers—is to go ahead, leaving the rest of the group behind to follow them. Those who go first will leave signs for the others, so that these will know which path to take. Thus, the speakers here are excluding the hearers.

116. *ele goʔjhuŋ goʔjhuŋ daru=ki=te eŋɖi buŋ toʔba=ga*
 1PL.EXCL path REP tree=PL=OBL pole INST pound.with.stick=FOC

thonɣ=ga co=na=le culha cawka=ki=te
 pound.with.rock=FOC go=MID.IRR=1PL.EXCL stove ECHO=PL=OBL

melay=ga co=na=le hinte thonliŋ roʔ=ga
 leave=FOC go=MID.IRR=1PL.EXCL LOC milk spill=FOC

co=na=le ho=jeʔ=ki=ga yo=ta yota am=pe
 go=MID.IRR=1PL.EXCL that=SG.NHUM=PL=FOC see=CVB REP 2=2PL

ele=te quluj=e=pe.
 1PL.EXCL=OBL follow=ACT.IRR=2PL

‘*We* (EXCL) will pound sticks and rocks against trees along the way (to make marks), leave stoves behind and spill milk there (as signs). Seeing these things, follow *us*.’ [MT, 1:161]

The inclusive plural form is also generally used when speaking about the Kharia, even with non-Kharia: We recorded a number of spoken texts dealing with the traditional history of the Kharia and their legendary travels through South Asia to their present homeland. The form *aniŋ* ‘1PL. INCL’ was regularly used throughout, although the story was addressed to the author, a non-Kharia. This seems to be due to habit, as a Kharia will not often speak to non-Kharias in Kharia. Thus, when referring to “our history” in Kharia, it will be natural to use the inclusive form as in the following example:

117. *tama iŋ aniŋ=aʔ gotar=aʔ kahni batay=na*
 now 1SG 1PL.INCL=GEN clan=GEN story tell=INF

col=taʔjd=ijn.

go=MID.PROG=1SG

'Now I am going to tell the story of *our* clan.'

[AK, 1:1]

Third-person forms

There are two further distinctions in the third persons:

Anaphoric vs. unmarked

There are two sets of forms for the third persons, one based on *adi*, the other on the demonstratives (5.8). The difference between these two groups is as follows: Forms based on the demonstratives, such as *ho* 'that', can be used in anaphoric, cataphoric and deictic function whereas those based on *adi* can only be used anaphorically and refer back to the topic.¹⁸ Also, forms based on *adi* can only be used with reference to humans. In the dual and plural, *adi* is usually realized as *ad-*.

118. *ho=kaṛ loʔkha doʔh=oʔ ro loʔkha doʔd=na loʔdho*
 that=SG.HUM soil take=ACT.PST and soil take=INF after

i=ghay adi=yaʔ rup buŋ=ga konon konon murti
 what=way ANAPH=GEN FORM INST=FOC small REP statue

bay=oʔ.

make=ACT.PST

'He [= God] took soil and, after taking soil, He made small statues in (= with) *His [own]* form.'

[AK, 3:10]

These forms have traditionally been considered reflexive pronouns (e.g. Biligiri, 1965: 37; Malhotra, 1982: 63), and this would seem to fit in well with example (118) above. However, as the following examples show, although *adi* can refer back to the subject of the clause it appears in, with a reflexive interpretation, this is not necessarily so. Instead, *adi* merely refers back to the present human topic and can appear in any function in the clause, including that of subject.

¹⁸ Anaphoric pronouns are also found elsewhere in South Asia, e.g. in Mundari (Osada, 1992: 66f), in Sinhalese (Gair, 2003: 782) and in at least some Dardic languages (Bashir, 2003: 845; 858).

119. *souʔb se maha beʔt=dom, simra, moŋ maŋa=te daʔ*
 all ABL big son=3POSS Simra one cave=OBL water

kuy=oʔ. ađi uʔh=oʔ.
 find=ACT.PST ANAPH drink=ACT.PST

‘His eldest son, Simra, found water in a cave. *He* drank [it].’

[AK, I:22]

120. *beŋa ho beŋi=yaʔ su-su lutui=te kach=oʔ*
 boy that girl=GEN put.on-RDP clothing=OBL untie=ACT.PST

ro ađi su=yoʔ.
 and ANAPH put.on=ACT.PST

‘The boy took off (= untied) the clothing that the girl was wearing and *he* put it on [himself].’

[BB, 1:76]

In fact, even if *ađi* is not the subject of the clause, it can also refer back to a topical individual which is not the subject of the clause it appears in, as in the following two examples, where *ađiyaʔ* refers back to the topic (not given here):

121. *hin=aʔ ghaʔd muđu beriya dada=dom=kiyar=te*
 that=GEN PURP one.HUM two.people elder.brother=3POSS=DU=OBL

ađi=yaʔ sori sori paʔtopur=te=ga ho=ki melay
 ANAPH=GEN with REP Patna=OBL=FOC that=PL leave

tu=yoʔ=ki.

DPT=ACT.PST=PL

‘Therefore they left one or two elder brothers behind with *him* in Patna.’

[MT, 1:158]

122. *laʔ ađi=yaʔ tay konon bhai, beghma, ho=kaŋ*
 then ANAPH=GEN ABL small brother Beghma that=SG.HUM

oŋdor=oʔ. ro deł=ki ho daru boʔ=te.
 hear=ACT.PST and come=MD.PST that tree place=OBL

‘Then *his_i* younger brother, Beghma_{*j*}, *he_j* heard [him_{*i*}]. And [he_{*j*}] came to the tree.’

[AK, 1:30]

The unmarked proforms (*hokaŋ, hojeʔ*) are literally “unmarked” in this respect: In addition to their deictic and cataphoric functions, they can also be used anaphorically. In fact, they often refer anaphorically to the same

entity as *adi* within the space of a single sentence, as in the following example.

123. *ghaḍ* *adi=te* *baṛa kharab laʔ=ki.* *tay moṇ moṇ*
 therefore ANAPH=OBL big bad EMOT=Mid.PST then one REP
bhai=ki=yaʔ jimi dhoʔd=ta dhoʔdta ho=kaṛ *ʈhisa=na*
 brother=PL=GEN name take=CVB REP that=SG.HUM call.out=INF
māre=yoʔ.
 start=ACT.PST
 ‘Therefore, *he* became very sad. Then, taking the names of the
 brothers one by one, *he* began to call out.’ [AK, 1:27f.]

In the following example, the unmarked status of these proforms is especially clear. Here, the two instances of *ho=kaṛ* refer to different individuals. Note that the first occurrence of *ho=kaṛ* in fact refers back to the *brahman* who has just been mentioned:

124. *muda moṇ brahman,* *ho=kaṛ_i* *ho=kaṛ=aʔ_j* *daru=te*
 but one Brahman that=SG.HUM that=SG.HUM=GEN tree=OBL
taṇa-l j anew=te yo=yoʔ.
 hang-PTCP holy.thread=OBL see=ACT.PST
 ‘But a Brahman_i, *he_i* saw the holy cord which *he_j* had hung on the
 tree.’ [AK, 2:14]

Human / Non-human

The unmarked proforms of the third-person generally distinguish between human and non-human reference in the singular, and occasionally in the dual and plural as well. These proforms derive from a demonstrative such as *u* ‘this’, *ho* ‘that’ (medial) or *han* / *hin* ‘that’ (distal) plus a second element: The proforms for human reference make use of *kaṛ* ‘person’, which is quite common in compounds but whose use as a free morpheme is now largely obsolete, except in the more southerly Kharia-speaking regions. Proforms for non-human reference make use of *jeʔ*. *jeʔ* is also often found alone in this function, but does not appear to have any lexical meaning.

125. *u=jeʔ=ga* *heke* *manus* *jati=yaʔ*
 this=SG.NHUM=FOC QUAL.PRS man ethnic.group=GEN

kahani.

story

'This is the story of humanity.'

[AK, 3:17]

For an example of *ho=kaŋ* with human reference, see (123) and (124) above.

The [\pm human] distinction is occasionally extended to the non-singular, and forms such as *ho=jeʔ=ki* 'they' (NHUM) are encountered, as in the fourth line in (116) above and in the following two examples. In forms such as these, the gloss of *jeʔ* as 'SG.NHUM' for 'non-human, singular' is of course inappropriate but since these occurrences are rather rare, the most common function was chosen for the sake of providing a uniform gloss.

126. *akhir ho=ki[=te] pata la? guɖ=ki no ho=jeʔ=ki*
 finally that=PL=OBL KNOWN EMOT C:TEL= MID.PST Cmpl that=SG.NHUM=PL
mudui=ki=ya? kamu=ga aw=ki.
 enemy=PL=GEN work=FOC QUAL= MID.PST
 'In the end, they found out that *these* [deeds] were the work of the enemy.'
 [MT, 1:103]

127. *hin=a? thon=ga je bajhe=sikh=oʔ=may*
 that=GEN for=FOC CREL get.trapped=PERF=ACT.PST=3PL
ho=jeʔ=ki=jo jal tay oʔj=kon ob-leŋ goʔh=oʔj,
 that=SG.NHUM=PL=ADD net ABL take.out=SEQ CAUS-fly C:TEL=ACT.PST.1SG
melay goʔh=oʔj.
 leave C:TEL=ACT.PST.1SG
 'Therefore, I also took from the net *those* (i.e., the birds) which had been trapped and released them (= let them fly, left them).'
 [BB, 2:44]

jeʔ need not appear together with a demonstrative. Consider the following example, where it appears with the quantifiers *sob* 'all' and *haʔdo* 'half':

128. *muda moŋ dhāgar kongher, ber merom gupa=na*
 but one servant boy who goat shepherd=INF
laʔ=ki, ho=kaŋ iɖib=te aɖi=ya? bāʔa kuda
 IPFV= MID.PST that=SG.HUM night=OBL ANAPH=GEN share millet

koloŋ=te sob=*je?* *um* *no?*=*na* *la?*=*ki* *ro*
 bread=OBL all=SG.NHUM NEG eat=INF IPFV=MID.PST and

ha?do=je? *un=na* *la?*=*ki*.
 half=SG.NHUM place=INF IPFV=MID.PST

‘But one servant boy, who used to tend the goats, he didn’t use to finish *all* his share of the millet bread at night and he placed [down] *half* of it.’ [BB, 1:13]

je? also appears in contexts in which it would seem to be translatable by ‘thing(s)’ or ‘matter(s)’, which may have been its original meaning:

129. *raja* *rajvara=ki=ya?* *je?* *um=pe* *koŋ=si?*.
 king ECHO=PL=GEN SG.NHUM NEG=2PL find.out=PERF
 ‘You don’t know the first thing about the king and his court’s matters.’

Finally, *je?* may be used without a demonstrative or quantifier in an attributive construction reminiscent of headless relative clauses (adapted from Biligiri, 1965: 78):

130. *taŋay=si?* *je?* ‘the one which has hung’
 hang=PERF SG.NHUM

del=ki *je?* ‘the one who came’
 come=MID.PST SG.NHUM

It might be best to consider these last two examples to be “relative clauses” in which *je?* is a proform modified by a predicative attribute. See Section 7.6 for a further discussion of “relative clauses”

The [+human] distinction is not strictly observed, and the “human”-referencing proforms are occasionally used in reference to non-human (and even inanimate) referents and *vice-versa*. This is not true of anaphoric *ađi*, however, which is restricted to human reference.

=*je?*

With non-human reference

131. *dhirom* *dhirom* ho=*je?*=*ga* *tama* *lebu=ki*
 slowly REP that=SG.NHUM=FOC now person=PL

“*hindustan*” *gam=te=ki*.

Hindustan say=ACT.PRS=PL

‘Gradually, the people today call *that* [= India] “Hindustan” ’

[MT, 1:55]

With human reference

The use of *je?* with human reference, although attested, is extremely rare.

Here one example is presented:

132. *.hin ti? tay* *u=je?=ki* *katna* *moskil* *se*
therefore this=SG.NHUM=PL how.much difficult ABL

katib *ḍom=taʔ*.

collect-<CAUS> PASS=ACT.PROG

therefore *they* [= the Kharia] are all being collected with great difficulty.’

[MS, 1:301]

=kaṭ

With human reference

133. *ho=ki* *soub* *pujapaṭh=a?* *ghaḍ* *ḍoko* *god=ki=may*,
that=PL all sacrifice=GEN for sit.down C:TEL=MID.PST=3PL

cerocagordi, *pahan*, *khariya* *pahan*, *ho=kaṭ*
on.all.four.sides priest Kharia priest that=SG.HUM

ḍel=ki, *pujapaṭh* *karay=na*.

come=MID.PST sacrifice do=INF

‘They all sat down for the sacrifice, on all four sides, the priest, the Kharia priest, *he* came, to do the sacrifice.’

[AK, 2:18]

With non-human reference

Unlike human reference with *je?*, non-human reference with *kaṭ* is quite common. In fact, it seems that *ho=kaṭ*, etc., are in the process of becoming general third-person proforms, completely unmarked for humanness or even animacy.

Reference to an animal

134. *ro* *ho=kaṭ=a?* *komaṇ=te* *bay=o?=ki*.
and that=SG.HUM=GEN meat=OBL make=ACT.PST=PL

‘And [they] cut off *his* [= the deer’s] meat. (lit.: ‘they made *his* meat’)’

[AK, 1:54]

Non-animate reference

Reference to past actions

135. u=kaṛ=aʔ thoŋ buṛha=ki=te iku²d iku²d
 this=SG.HUM=GEN for old.man=PL=OBL very REP
 dhain mane=na ayiʔj.
 thank honor=INF QUAL.PRS
 'For this [= for what they did], the elders should be thanked and honored (= it is to thank [and] honour the elders).' [MT, 1:183]

Reference to places

136. ho hisab se ho=kaṛ=aʔ jimi khaybar
 that manner INST that=SG.HUM=GEN name Khyber
 ghaṭi hoy=ki.
 little.valley become=MID.PST
 'In that way *its* name became Khyber Pass.' [MS, 1:52]

137. moŋ bandho aw=ki. ho=kaṛ=a[ʔ] siniŋ=te moŋ
 one pond QUAL=MID.PST that=SG.HUM=GEN side=OBL one
 kulu aw=na laʔ=ki.
 turtle live=INF IPFV=MID.PST
 'There was a pond. On *its* side there lived a turtle.' [TK, 1:1]

Finally, *kaṛ* is used in other expressions denoting human beings, in line with its original meaning 'person'. In these cases it will be considered a free contentive morpheme here, although its exact status awaits further study.

138. uʔphe kaṛ '(the) third person' gupa kaṛ 'shepherd'
 three person guard person
 likha kaṛ 'author'
 write person
 139. cuṭa²d kaṛ cuṭa²d kaṛ konon=ga aw=ki, saŋgod=na
 small person REP small=FOC QUAL=MID.PST walk=INF
 um pal=na laʔ=ki.
 NEG be.able=INF IPFV=MID.PST
 'The child (= *small person*) was much too small, it couldn't walk.'
 [MT, 1:157]

140. *ḍoriyay kaṛ=ko mohto heke la!*
 lead PERSON=CNTR *mohto* QUAL.PRS VOC
 ‘The leader is a *mohto*!’ [Kerkeṭṭā, 1990: 2]

5.7 INTERROGATIVES, INDEFINITES AND NEGATIVE INDEFINITES

For the sake of presentation, in this section interrogatives in general are discussed. As indefinite proforms and negative indefinites are largely based on interrogative forms, they will also be dealt with here.

Interrogatives

There are seven basic interrogative forms:

- | | |
|----------------------------|--|
| 141. Free morphemes: | Bound morpheme: |
| <i>ata</i> ‘what?, which?’ | <i>ber, behar</i> ¹⁹ ‘who?’ |
| <i>atu</i> ‘where?’ | <i>a=</i> ‘Q’ |
| <i>ina</i> ‘why?’ | <i>i</i> ‘what?’ |
| | <i>kiṛte, kiṛjte</i> ‘how many?’ |

a= and *i* combine with free morphemes to derive other interrogatives. With a number of contentive morphemes, such as *ghay* ‘way’, *i* forms a phonological word with the typical LH pitch pattern (2.5), but not with others, such as *bhere* ‘time’ *a=* always combines with some element to form a phonological word, including the postpositions *boṛtay* and *khoṛtay* ‘up to’ and the oblique marker *=te*.

- | | |
|--|---|
| 142. <i>a=bo?</i> ‘where?’ (<i>bo?</i> ‘place’) | <i>i bhere</i> ‘when?’ (<i>bhere</i> ‘time’) |
| <i>a=te</i> ‘where?’ (<i>=te</i> ‘OBL’) | <i>i=ghay</i> ‘how?’ (<i>ghay</i> ‘way’) |
| <i>a=kho?</i> ‘where?’ (cf. <i>kho?</i> ‘place’) | <i>i=te</i> ‘where?’ (<i>=te</i> ‘OBL’) |
| <i>a=tiṛj</i> ‘to where?’ | |
| <i>a=khoṛtay, a=boṛtay</i> ‘how far, up to where?’ | |

¹⁹ The form *behar* seems to be a combination of the original Kharia form *ber* and the Sadri marker for third person inalienable possession, *=har*, modeled along the lines of the forms in Sadri. Cf. Sadri *ke / ke=har* ‘who?’, with no apparent difference in meaning. In other words, **ber=har > behar* in Kharia. The origins of this form in Sadri are unclear, and it may ultimately derive from Munda there, although it would seem reasonable in our case to assume that the Kharia form is based on the Sadri form, especially since the Sadri possessive marker *=har* is used.

As noted in Section 4.2, interrogatives may also be used predicatively, provided that they are not case-marked:

143. *am* *i=yo³b?* 'What did you do?'
 2sg what=ACT.PST.2sg
144. *am=te* *i=ki?* 'What happened to you?'
 2sg=OBL what=MD.PST
145. *bero=te=ga* *tiri³b* *ḡal=e* *la?* *i=ghay=na?*
 sun=OBL=FOC cloud cover=ACT.IRR then what=way=MD.IRR
 'If a cloud covers the sun, then *how will it be?*' [Kerkeṭṭā, 1990: 5]

Indefinites

Some interrogatives, such as *atu* 'where', combine with the additive focus marker =*jo* 'also' to serve as indefinites, e.g. *atu=jo* 'wherever'. For some indefinites, however, there is no corresponding interrogative, such as *jahā* 'something' or *jahāy(=ga)* 'someone; whoever'. *jahā* may also be used attributively. The most common indefinites are:

146. *jahāy, jahāy=ga, ber=jo* 'someone; whoever'
 jahā, i=jo 'some(thing); whatever'
- atu=jo* 'somewhere; wherever'
- jahā bhere* 'sometime; whenever'

Complex interrogatives, such as *i bhere* 'when', generally form indefinites by repeating the interrogative element, in this case *i*, separated by *no* 'or', e.g., *i no i bhere* 'whenever', literally 'what or what time'. With *muḡu* 'one', *la?* 'then' is used instead of *no*.

The most common indefinites are given in Table 5.5, with their corresponding interrogatives. Following this, a few examples illustrating their use are given.

Table 5.5: Indefinite pronominals and their corresponding interrogatives

	<i>Indefinite</i>	<i>Corresponding interrogative</i>
<i>Temporal</i>		
‘ever’	<i>kehiyo(=jo), kahiyo(=jo)</i>	–
‘whenever’	<i>jahā bhere</i> <i>i no i bhere</i>	– <i>i bhere</i>
‘since whenever’	<i>i no i bhere tay</i>	<i>i bhere tay</i>
‘up to whenever’	<i>i no i bhere boʔtay</i>	<i>i bhere boʔtay</i>
<i>Locative</i>		
‘wherever’	<i>atu=jo, a=te=jo</i>	<i>atu, a=te</i>
‘from wherever’	<i>atu no atu tay</i>	<i>atu tay</i>
‘to wherever’	<i>a=tiʔj=jo</i> <i>abtu no abtu²⁰</i>	<i>a=tiʔj</i> –
<i>Other</i>		
‘however’	<i>i=ghay no i=ghay</i>	<i>i=ghay</i>
‘however many’	<i>kiʔte no kiʔte, kiʔjte no kiʔjte</i>	<i>kiʔte, kiʔjte</i>
‘some . . . or other’	<i>ata no ata, i no i,</i> <i>muḍu laʔ muḍu</i>	<i>ata, i, –</i>
‘whoever’	<i>jahāy, jahāy=ga,</i> <i>ber=jo</i>	– <i>ber</i>
‘whatever’; ‘something’	<i>i no i, keʔko, jahā</i>	<i>i, –, –</i>

147. *maha=ki goʔjhuŋ hinte i no i jait=yaʔ*
 big=PL path LOC what or what ethnic.group=GEN

lebu=ki=yaʔ peʔ nokh=oʔ=ki ro konon
 person=PL=GEN rice eat=ACT.PST=PL and small

goʔd=ki=may.

C:TEL=MID.PST=3PL

‘The elder ones ate the food (= rice) of people of *various* (= *some or other*) tribes along the way and [thus] became impure (= small).’

[MT, 1:179]

²⁰ Structurally speaking, *abtu* appears to derive from *atu* ‘where’ and the causative infix *-b-*. If so, then the details of its development are unclear at present.

148. *jhari po?da hinte muḍu la? muḍu gāysi*
 all village LOC one.HUM then one.HUM spy

aw=ta=ki=ga.

QUAL=MID.PRS=PL=FOC

‘In all villages there are [always] *some* spies.’ [MT, 1:207]

149. *hin=a? thoṅ ubhroṅ=ko jahāy=ga ho? bay=te=may,*
 that=GEN for nowadays=CNTR INDEF.HUM=FOC house build=ACT.PRS=3PL

la? simēt, ṭṭā roke?ḍ ro soreṅ=a? ho?
 then cement brick sand and stone=GEN house

bay=te=ki.

build=ACT.PRS=PL

‘Therefore, when *someone* makes a house nowadays, then they build a house of cement, brick, sand and stone.’ [AK, 5:10]

Negative indefinites

When indefinites are negated, the resultant meaning is ‘no one’, ‘nowhere’, ‘nothing’, etc. When the clause is negated, the /o/ in =*jo* is generally deleted before the /u/ in *um* ‘NEG’:

150. *ber=jumbo?* ‘no one’ (< *ber=jo umbo?*)
 who=ADD NEG

151. *atu=jum=ijn col=ki.* ‘I didn’t go anywhere.’
 where=ADD.NEG=1SG go=MID.PST

In *ber=jo*, the interrogative component *ber* ‘who’ can even be omitted.

152. (*ber*)=*jumay col=ki.* ‘No one went.’ (< *ber=jo umay col=ki*)
 who=ADD.NEG.3PL go=MID.PST

When *um* does not directly follow *jo*, there are no further changes:

153. *ijn kehiyo=jo ho=te um=ijn col=ta.*
 1SG ever=ADD that=OBL(= ‘there’) NEG=1SG go=MID.PRS
 ‘I never go there.’

154. *pheinga=te jo?=na ghađ atu=jo kořom ula?*
 grasshopper=OBL eat=INF PURP where=ADD soft leaf
um bańce=ki.
 NEG be.saved=MID.PST
 ‘Nowhere were soft leaves left over (= saved) for the grasshopper to eat.’ [TK, 2:30]

At least in the written language, the fusion of /o/ and /u/ is not obligatory:

155. *gam=na lam=o? muda i=jo um gam=o?*
 say=INF want=ACT.PST but what=ADD NEG say=ACT.PST
 ‘He wanted to speak but didn’t say anything.’ [TK, 1:56]

Note that case-marked indefinites (including the genitive) place the case marker before the additive focus marker *jo*:

156. *iđib=te solo? bhabru=te muda ber=a?=jo lemeđ*
 night=OBL dog bark=ACT.PRS but who=GEN=ADD sleep
um pi?j=ta, sob aram se lemeđ=ta=ki.
 NEG break=MID.PRS all comfort INST sleep=MID.PRS=PL
 ‘At night the dog barks, but nobody wakes up (= the sleep of nobody breaks), all sleep comfortably.’ [BB, 3:65]

An alternative means of expressing ‘nothing’, ‘no one’, etc. is by using a quantifier denoting totality and negating the clause:

157. *lekin soub baru-bo?=ga aw=ta=may, soub umay*
 but all good-INTENS=FOC live=MID.PRS=3PL all NEG.3PL
kole?j=ta.
 fight=MID.PRS
 ‘But all live together very well, no one fights (= all do not fight).’ [AK, 5:32]

5.8 DEMONSTRATIVES

Kharia has a three-way demonstrative system, with some evidence for an earlier four-way system. Table 5.6 presents an overview.

Table 5.6: The demonstratives

<i>Proximal</i>	<i>Medial</i>	<i>Distal</i>
<i>u</i>	<i>ho</i>	<i>hin, han, se</i>

hin and *han* differ little in meaning in actual usage, if at all. Along with Malhotra (1982: 53), we can find no evidence for Biligiri's (1965: 65) analysis of *han* as restricted to visible and *hin* to non-visible entities. However, unlike Malhotra, we also find no evidence for *hin* and *han* being in complementary distribution. Instead, our data indicate that they are in free distribution, with *hin* being the more common of the two.

It is likely that there once was a difference between *han* and *hin* (and perhaps still is for some speakers) and that Biligiri's analysis held for his speaker at that time, or that the two demonstratives differed in some other, similar way: When questioned during interviews as to the difference between *hin* and *han*, some speakers indicated that *hin* refers to an entity which is farther away than one marked by *han*. This fits in well with Biligiri's analysis of *hin* as being restricted to non-visible, or perhaps very distant, entities. However, despite this, there is no evidence in our data for such a distinction in actual speech, so that we conclude that the two are now freely interchangeable.

se, a borrowing from Sadri, is only seldomly encountered and is generally only found with temporal, less often with locative or other adverbials, cf. *se bhere* 'that time', *se dinu* 'that day', *se bo?=te* 'at that place', *se lekhe* 'like that'

These demonstratives may appear at the beginning of the semantic base as a determiner, preceding all other constituents, or with the exception of *se* they may combine with *kaɾ* 'person' and *je?* 'SG.NHUM' to form "pronominals" (5.6) consisting of a single phonological word. Again with the exception of *se*, they may also combine with the oblique marker =*te* to form locative adverbials, with *ghay* 'way' to form manner adverbials or with *ti?j* 'side, direction' to form goal-directed locative adverbials, again as single phonological words.

158. <i>hin bhere</i> 'at that time' that time	<i>u=te</i> 'here' this=OBL
<i>ho=te</i> 'there' (MEDIAL) that=OBL	<i>han=te, hin=te</i> 'there' (DISTAL) that=OBL

159. *u=kaṛ* 's/he (here)', *ho=kaṛ* 's/he (MEDIAL)
 this=SG.HUM that=SG.HUM
hin=kaṛ, *han=kaṛ* 's/he (DISTAL)'
 that=SG.HUM

160. *ho=ghay*
 that=way
 'that way, like that'

161. *u* *goṭa* *duniya=te*
 this entire world=OBL
 'in this entire world'

162. *han=tiʔj* *u=tiʔj*
 that=side this=side
 'this way and that, in this direction and that'

The demonstrative, if present, is always the first element of the semantic base, except in the case of heavy constituents such as preposed attributive clauses, where it generally follows the attributive clause.

163. *tay* *adj=te=jo* *batay=oʔ* *no* *dhāy[=e]*
 then ANAPH=OBL=ADD tell=ACT.PST Cmpl hurry=ACT.IRR
am, *han* *maṛa* *boʔ=te* *ḍa[ʔ]* *[a]iʔj*.
 2SG that cave place=OBL water QUAL.PRS
 'Then he said to him, too, "Hurry, *at that cave* (distal) there is water."' [AK, I:38]

164. *ro* *tama* *am=pe* *u* *nāw* *kuṭum=te=ga* *sadi*
 and now 2=2PL this nine family=OBL=FOC marry
biha *kersonj=na=pe*
 marry marry=MID.IRR=2PL
 'And you will marry *in only these nine families* [AK, I:69]

As noted, the demonstrative generally follows attributive clauses and precedes all other elements of the semantic base. It can, however, precede these "heavy" constituents, although speakers we worked with, while they accepted this use, preferred examples in which the demonstrative follows the "heavy" propositional attribute. In the following an example for both

the preferred order as well as the alternate, possible but less common order is presented:

Preferred

165. *ho=ki eboʔ=na laʔ=ki=may* ho bhere *ijn*
 that=PL play=INF IPFV=MID.PST=3PL that time 1SG
sahar=te aw=ki=jn
 city=OBL QUAL=MID.PST=1SG
 'I was in the city *at the time they were playing.*'

Acceptable, not common

166. ho *ho=ki eboʔ=na laʔ=ki=may* bhere *ijn*
 that that=PL play=INF IPFV=ACT.PST=3PL time 1SG
sahar=te aw=ki=jn
 city=OBL QUAL=ACT.PST=1SG
 'I was in the city *at the time they were playing.*'

As noted in Section 5.6, the demonstratives may also be used anaphorically and cataphorically, in addition to their deictic use. In general, *ho* refers to something which has already been said, or to the first of two or more entities already mentioned, whereas *u* refers to something which is about to be said, or to the last of two or more entities already mentioned.

Anaphoric

167. *laʔ porha=jo col=ki ro yo=te,* ho maʔa boʔ=te
 then Porha=ADD go=MID.PST and see=ACT.PRS that cave place=OBL
dam=ki daʔ=te yo=yoʔ.
 arrive=MID.PST water=OBL see=ACT.PST
 'Then Porha also went and looks, [he] arrived *at that cave* [i.e., which I just mentioned] [and] saw the water.' [AK, I:39]

Cataphoric

168. *u=jəʔ onʔor=e!*
 this=SG.NHUM hear=ACT.IRR
 'Listen to *this* [i.e., to what I'm about to say]!'

hin and *han* may also be used anaphorically. *hin* is especially common in correlative constructions (7.6.1):

169. ata bhere *bulbul* *po?da* *raja* *nāwkod* *najor*
 CREL time Babylon village king Nawkod Najor
- israeli* *lebu=ki=te* *bādi* *bay=kon* *misar* *raj*
 Israeli person=PL=OBL imprisonment make=SEQ Egypt kingdom
- ḏo?ḏ=na* *la?=ki* *hin* bhere *ho=ki* *purkha=ki*
 take=INF IPFV=MID.PST that time that=PL ancestor=PL
- khaṛiya* *buṅ=ga* *aw=ki=may*.
 Khaia INST=FOC live=MID.PST=3PL
- ‘When Nawkod Najor, king of the village of Babylon, imprisoned the Israelis and was taking them to Egypt, *at that time* these ancestors [of the Israelis] lived with the Kharia.’ [MT, I:14]

170. a=te ate *bes* *bes* *ṭhāro* *kuy=o?=ki*,
 Q=OBL(=‘where (CREL)’) REP good REP place find=ACT.PST=PL
- hin=te* *hinte* *khaṛiya* *go?jlo?* *kamu=na*
 that=OBL(=‘there’) REP Kharia ricefield work=INF
- la?=ki=may*.
 IPFV=MID.PST=3PL
- ‘Wherever they found good land (= places), *there* the Kharia would work the rice fields.’ [MT, 1:8]

se ‘that’, from Sadri, is used somewhat differently. It is found predominantly in temporal and locative expressions ((171)–(172)), and very commonly in correlative constructions (173). However, it may not be used in reference to animates or concrete objects:

171. *se* *bhere* ‘then, at that time’
 that time
172. *se* *bo?=te* ‘there, at that place’
 that place=OBL

173. *ghaḍ aḍi je bhere eṅ=ki r[o] aba=qom=te*
 therefore ANAPH CREL time return=MID.PST and father=3POSS=OBL
souʔb bhai=kiyar=aʔ soriʔ poʔom=te kholay kholay
 all brother=HON=GEN together bundle=OBL open REP
ob-yo=na laʔ=ki=may, se bhere aḍi=yaʔ poʔom=te
 CAUS-see=INF IPFV=MID.PST=3PL that time ANAPH=GEN bundle=OBL
soreṅ kui=ki.
 stone find=MID.PST
 ‘Therefore, *when* he returned, and along with all his brothers they opened the bundles and were showing [them] to their father, *at that time* he found a stone in his own bundle.’ [AK, 1:73]

5.9 QUANTIFIERS AND CLASSIFIERS

5.9.1 Quantifiers

Quantifiers in Kharia appear in general before their lexical head:

174. *souʔb lebu=ki=yaʔ thoṅ* ‘for all people’
 all person=PL=GEN for
175. *jhari khariya=ki* ‘all Kharia’
 all Kharia=PL
176. *kaʔiʔj deri* ‘some time’
 some time

As noted in Section 5.4, if a quantifier denoting non-singularity is present, the semantic base is often unmarked for number. Cf. the following examples:

177. *ro souʔb bhai=ga ho=ghay=ga enem ḍa[ʔ]*
 and all brother=FOC that=way=FOC without water
uḍ=ga eṅ ḍel=ki=may.
 drink=FOC return come=MID.PST=3PL
 ‘And *all brothers* thus returned, without drinking water.’

[AK, 1:51]

178. *bajhe konthed=te jughay kole?, heirla, perwã,*
 get.trapped bird=OBL many parrot green.pigeon pigeon

kurkur, mayna, ro kinir siṅkoy aw=na
 dove maina.bird and forest chicken QUAL=INF

laʔ=ki=may.

IPFV=MID.PST=3PL

‘Among the birds caught there were *many* parrots, green pigeons, other pigeons, doves, *maina* birds and wild chickens.’ [BB, 2:12]

The only postposed quantifiers which we are aware of are *geinka* ‘every’ and *bheir* ‘entire’, both loans from Sadri:

179. *khawna=ko pethiya geinka ol kay=na*
 a.kind.of.sweets=CNTR market every bring BEN=INF

laʔ=ki=kiyar.

IPFV=MID.PST=DU

‘And [at] *every market* they would bring sweets for him.’ [RD, 2:4]

180. *musniṅ dinu bheir lam=oʔ=ki muda umay*
 one.day day entire search=ACT.PST=PL but NEG.3PL

kuy=oʔ.

find=ACT.PST

‘One day, they searched *all day* but didn’t find [the demon].’

[MT, 1:68]

Kharia Numerals

There are two sets of numerals in Kharia. One, of Kharia origin, is no longer in general use. Many speakers are, however, familiar with these numerals, although the data they provide shows a great deal of variation. Despite the fact that these numerals are no longer actively used, as speakers who this author worked with consciously tried to use “pure” Kharia, these numerals are often found in texts stemming from these speakers. Also, they are very often found in Pinnow’s (1965a, b) texts.

The other set has been borrowed from Sadri. Only the Sadri numerals are at all common, and even the lower numerals in actual speech are all borrowed from Sadri. Occasionally, English numerals are also used. Numerals of Sadri origin generally occur with numeral classifiers, although this is not obligatory. If the Kharia numerals are used, the classifiers are

only very seldomly used, at least by modern speakers. In the texts in Pinnow (1965), however, the Kharia numerals are occasionally found in combination with numeral classifiers. Thus, the fact that the classifiers are seldom used with Kharia numerals is perhaps due to the unfamiliarity of modern speakers with the Kharia numerals.

Cardinal numbers

There is great deal of uncertainty concerning the numerals of Kharia origin, and only the forms for 1–3, 6 and 10 are relatively uniform for all speakers. Other numerals can vary greatly from one speaker to another. In Table 5.7, beginning with '4', the first numeral given is the most commonly cited form, followed by less commonly cited variants.

Table 5.7: The Kharia numerals from 1 to 10

1	<i>muḍu</i> (HUM), <i>moŋ</i> (NHUM)	6	<i>tibru</i> , <i>tibhru</i> , <i>ʃibru</i>
2	<i>ubar</i>	7	<i>tham</i> , <i>thom</i> , <i>thoŋ</i> , <i>ghul</i>
3	<i>uʔphe</i>	8	<i>thom</i> , <i>ghal</i> , <i>tham</i> , <i>thomsinʃ</i>
4	<i>iʔphon</i> , <i>tham</i>	9	<i>thomsinʃ</i> , <i>ghal</i> , <i>tomsinʃ</i> , <i>ghul</i>
5	<i>moloy</i> , <i>thum</i>	10	<i>ghol</i>

Note that the numeral '1' has two forms, one used with humans, *muḍu*, and one which is unmarked, *moŋ*. Thus, in a sense these two forms have a kind of "built-in" classifier. They will be glossed as 'one.HUM' and 'one', respectively.

There are two systems for the numerals from 11 to 19, presented in Table 5.8. The first simply combines the word for '10' with the respective single digit. There is also another system in which all numerals from 11 to 19 have their own designation. This system is given to the right of each column in Table 5.8. The data for this second group were provided by a group of young men (all aged 17), who read them to the author from their school notes. All confirmed that they had been taught these numerals in school. However, they themselves use the numerals borrowed from Sadri.

Table 5.8: The Kharia numerals from 11 to 19

11	<i>ghol moŋ</i>	<i>ghul</i>	16	<i>ghol tibru</i>	<i>rabeʔ</i>
12	<i>ghol ubar</i>	<i>gholsinʃ</i>	17	<i>ghol tham</i>	<i>tarsiʔ</i>
13	<i>ghol uʔphe</i>	<i>ʃak</i>	18	<i>ghol thom</i>	<i>qabayeʔj</i>
14	<i>ghol iʔphon</i>	<i>ʃḍya</i>	19	<i>ghol thomsinʃ</i>	<i>qubki</i>
15	<i>ghol moloy</i>	<i>rabaʔ</i>			

One elderly speaker who the author consulted on these numerals, however, who has been teaching Kharia at the university for a number of years, was unfamiliar with these forms and suggested that they have perhaps recently been formed in an attempt to give Kharia “pure” numerals. In the absence of any evidence for or against this view, this topic will not be pursued further here.

The designation for ‘20’ is *ekṛi* (< Indo-Aryan, cf. Bengali *kuṛi* ‘a score’ Sadri *kori* ‘a score’). Pinnow (1965a: 6) also cites *ukṛi*. For 21–29, the single digits are added to *ekṛi*, e.g., *ekṛi moṇ* ‘21’, etc.

There are no simple words in the Kharia numeral system for ‘30’, ‘40’, etc. There are two different strategies for expressing numbers above ‘29’ These are summarized in Table 5.9.

- The first system, the vigesimal system, counts in scores and adds the appropriate numeral from ‘1’ to ‘19’ to the correct multiplication of ‘20’
- The second system counts in decades, with multiples of ‘10’ plus the respective single digit. The fact that this system has no decadal term for ‘100’ suggests that it is the younger of the two.

We are not aware of any numerals of Kharia origin higher than 100.

Table 5.9: Kharia numerals from 30 to 100

<i>Vigesimal system</i>	<i>Decadal system</i>
30 (<i>moṇ</i>) <i>ekṛi ghol</i> , literally ‘(1–)20–10’	<i>uṛphe ghol</i> ‘3–10’
31 (<i>moṇ</i>) <i>ekṛi ghol moṇ</i> , etc.	<i>uṛphe ghol moṇ</i> , etc.
40 <i>ubar ekṛi</i> ‘2–20’	<i>iṛphon ghol</i> ‘4–10’
50 <i>ubar ekṛi ghol</i> ‘2–20–10’	<i>moloy ghol</i> ‘5–10’
60 <i>uṛphe ekṛi</i> ‘3–20’	<i>tibru ghol</i> ‘6–10’
70 <i>uṛphe ekṛi ghol</i> ‘3–20–10’	<i>tham ghol</i> ‘7–10’
80 <i>iṛphon ekṛi</i> ‘4–20’	<i>thom ghol</i> ‘8–10’
90 <i>iṛphon ekṛi ghol</i> ‘4–20–10’	<i>thomsinṅ ghol</i> ‘9–10’
100 <i>moloy ekṛi</i> ‘5–20’	<i>moloy ekṛi</i> ‘5–20’

Ordinals

Information on Kharia ordinals is very scant. The following discussion is based on what little evidence could be culled from forms in the texts in Pinnow (1965a) and Kerkeṭṭā (1990).

The native term for 'first' is generally *seŋ*, which is quite common:

181. *porob=a?* *musniŋ*, *seŋ* *dino=te* *o?*=*ya?* *gomke*
 feast=GEN one.day first day=OBL house=GEN lord

kan=te.

fast=ACT.PRS

'On one day of the feast, *the first day*, the lord of the house fasts.'

[HJPa:71]

Alternatively, the genitive of *absi^{2b}* 'beginning' can also be used. The following is the only example of this construction in our data, from Pinnow (1965a: 90):

182. *absib=a?* *eto?*
 beginning=GEN order

'the first commandment', literally 'the commandment of the beginning'

Other ordinals are derived as follows: In reference to humans, *kaŋ* follows the corresponding cardinal number, at least in the texts in Kerkeŋtā (1990):

183. *u?phe kaŋ* '(the) third person' *i?phon kaŋ* '(the) fourth person'
 three person four person

For non-humans / inanimates, there are three strategies: First, the respective cardinal number is placed in the genitive:

184. *ubar=a?* *etoŋ* '(the) second commandment'²¹
 two=GEN order

In a second strategy, found in the texts in Kerkeŋtā (1990), beginning with *ubar* 'two' the compound form of *absi^{2b}* 'beginning', *-si^{2b}*, is used to form ordinals, e.g. *i?phon-si^{2b} yona* 'fourth scene', etc. Note that the *u* in *ubar* is dropped: *bar-si^{2b} āk* 'second act (of a play)'

²¹ It should be noted that all the examples given here referring to the Ten Commandments derive from the same source, a Kharia translation of Martin Luther's *Kleiner Katechismus* (given in Pinnow, 1965a: 90ff.). These should be viewed with caution, as in this text the exceptional form for 'first', *absib=a?*, given above, is also found.

In a third strategy, also found in the texts in Kerkeṭṭā (1990), cardinal numbers are used, with no special marking, similar to the strategy in (183): *ubar ebo?* ‘second act / scene’, *iṭphon yona* ‘fourth scene’

Finally, Malhotra (1982: 126) writes of yet another strategy for ordinals beginning with ‘second’ which is not found in our corpus. She notes that ordinals above ‘first’ are formed by means of postpositions meaning ‘after’ or ‘behind’ following the next lower numeral. Her two examples (adapted):

185. *boriya=? loṭḍho* ‘third’ *moloy=a? kundab* ‘sixth’
 both=GEN after five=GEN behind

Unfortunately, as these forms are no longer used, no further comments can be given here. Nowadays, only the Sadri forms, given below, are used.

Distributives

Unfortunately, information on native distributive terms is even scantier than that on native ordinals. In fact, only two such terms, both from Pinnow (1965a), could be located so far:

186. *mo<ne>ṅ* ‘one each’ *u<nu>ṭphe* ‘three each’

Note that both of these terms employ the native cardinal number and the *-NV-* infix, which is slightly irregular in *mo<ne>ṅ*, as the *-V-* in *-NV-* otherwise always has the same quality as the vowel of the underived morpheme. Although it is highly tempting to assume that this was once a productive means of deriving distributives in Kharia, in the absence of conclusive evidence this must remain speculative, especially since both forms derive from the same text (see HṆṆa: 146).

Indo-Aryan Numerals (from Sadri)

Cardinal numbers

As mentioned above, only the numerals which have been borrowed from Sadri are in common use. Occasionally, the Hindi numerals, which are almost identical to these, are employed. Table 5.10 presents the lower numerals and a few of the higher numerals.

Table 5.10: Numerals borrowed from Sadri

0 <i>sun</i>	6 <i>chaw</i>
1 <i>ek</i>	7 <i>sat</i>
2 <i>dui</i>	8 <i>aṭh</i>
3 <i>tin</i>	9 <i>naw, nāw</i>
4 <i>cair, ceir</i>	10 <i>das</i>
5 <i>pāc</i>	

Others: 100 *say, saw, sos*, 1,000 *hajar*, 100,000 *lakh*

Ordinals

The ordinals have also all been borrowed from Indo-Aryan. With the exception of 'first', all are pronounced as in Hindi and will not be presented here:

187. 1st *pohila*, 2nd *ḍusra*, 3rd *tisra*,

Collectives

With the exception of *bariya*, which is of Kharia origin, collectives are formed by adding =*o* directly to the Sadri numeral. This =*o* is probably identical with the homophonous classifier =*o* (5.9.2), although it may be related to the Hindi oblique plural suffix -*ḍ*. Table 5.11 presents the lower collectives.

Table 5.11: Collectives

<i>bariya, beriya, boriya</i> 'both'	<i>chaw=ō</i> 'all six'
(cf. <i>ubar</i> 'two')	<i>sat=ō</i> 'all seven'
<i>tin=ō</i> 'all three'	<i>aṭh=ō</i> 'all eight'
<i>cair=ō</i> 'all four'	<i>naw=ō</i> 'all nine'
<i>pāc=ō</i> 'all five'	<i>das=ō</i> 'all ten'

Approximatives

Approximation is expressed in Kharia by juxtaposing two or more cardinal numbers. Although the only examples in our corpus all involve the native Kharia numerals, it is reasonable to assume that the same strategy is employed with the cardinals from Sadri.

188. *la?* *kongher* *moŋ* *ubar* *uʔphe* *kolon* *ob-gur*
 then boy one two three bread CAUS-fall

kay=na *la?=ki*.

BEN=INF IPFV=MID.PST

'Then the boy would drop *a couple* (= 1, 2, 3) [*loaves of*] bread for them.'

[BB, 1:27]

189. ...*no biha olday hoy=ta beṭi uʔphe iʔphon*
 Cmpl marriage marriage become=MID.PRS girl three four

memon jou=te beṭa beṭi bariya=ga moṇ osel
 year up.to=OBL boy girl both=FOC one white

kokro siṅkoy giṛiṇ puja joʔ=te=kiyar.
 rooster fowl heat.of.sun worship eat=ACT.PRS=DU

that, the girl being married *for up to 3 or 4 years*, the boy and girl both sacrifice a white rooster to the Sun-God and eat [it].’

[HJPa:143, 1a]

5.9.2 Classifiers

There is a relatively small number of optional classifiers in Kharia. These are generally found only with cardinal numerals and, rarely, with quantifiers such as *kai* ‘several, a few’. In the following all of the classifiers which have been found in the data are presented, in alphabetical order. With the possible exception of the general classifier =*o*, all of the following classifiers are phonological words and can be separated from the element they follow, e.g., by the enclitic focal particle =*ga*:

190. *ek=ga ṭho=ga* ‘one (and only one)’
 one=FOC CLASS=FOC

cepu? ‘handful’

One occurrence in Pinnow’s (1965a) texts of the morpheme *cepu?* ‘handful’ is found, which appears to be used as a classifier, although this is uncertain:

191. *je? moṇ cepu? guḍray=o? ro do[t]h=o?*
 SG.NHUM one handful collect=ACT.PST and take=ACT.PST

ho beṭa=ga ro ḍuruṇ=o? ro ḍoḍ kay=o?

that boy=FOC and husk=ACT.PST and bring BEN=ACT.PST

‘He collected it, *one handful*, and that boy took [it] and husked [it] and brought [it back].’

[HJPa:136, no. 17]

gaṭ

gaṭ occurs three times in our corpus, all from the same southern speaker and each time found with human reference. It presumably derives from *kaṭ* ‘person’. Consider the following example, where the speaker begins

with the usual word for 'nine', *naw*, from Sadri, and the more common classifier *ʃho* (see below), generally not used with reference to humans. He then "corrects" this to the native form *thomsinj gaṛ*. Following this, he maintains the form *thomsinj* when referring to the daughters, but leaves out the classifier.

192. *hondo buṛha=ya? naw ʃho, thomsinj gaṛ gam=te=niṅ,*
 Hondo old.man=GEN nine CLASS nine CLASS say=ACT.PRS=1PL.INCL
be²d=d̄om, thomsinj bij=d̄om=ki aw=ki=may.
 son=3POSS nine daughter=3POSS=PL QUAL=MID.PST=3PL
 'The old man Hondo had *nine* sons, we say *thomsinj* [for 9], and 9 daughters.'
 [MS, 2:1]

go

This classifier is somewhat of a mystery. Malhotra (1982: 67) writes of it "go is the general numeral classifier, co-occurring with Indo-Aryan numerals." Her two examples are given in the following (slightly adapted):

193. *tin go kunṛu* 'three children'
 three CLASS child
 194. *pāc go kuyu* 'five pitchers'
 five CLASS pitcher

As there are no examples of this classifier in our data, nothing more can be said about it here other than to note that it is certainly not the general numeral classifier in the dialect of the speakers we worked with. It is also not attested in Pinnow (1965a), at least in the first 143 pages.

jan / jhan

This classifier is used exclusively with human reference. It can also be replaced by =o (see below). *j(h)an* derives from the Sadri classifier *-jan / -jhan*.

195. *tin jhan lebu=ki* 'three people' *tin jhan bhai=ki* 'three brothers'

j(h)an also attaches to names to denote the person who is explicitly mentioned as well as either those accompanying him or her at that moment or his or her family. Thus, it is not restricted to following a quantifier of some type, unlike most classifiers.

196. *birsi jhan* ‘Birsi and her friends (who are with her now)’
197. *aʔb=ko modi jhan moloy kulam hoy=ki=may.*
 now=CNTR Modi CLASS five brother become=MID.PST=3PL
 ‘But now Modi had four brothers (= but now, *Modi and his kin*
 became five brothers).’ [RD, 2:15]
198. *etwa=kiyar boriya=ga buḥha buḥhi modi jhan=te*
 Etwa=DU both=FOC old.man old.woman Modi CLASS=OBL
sewa karay=oʔ=kiyar.
 service do=ACT.PST=DU
 ‘Etwa and his wife, both of them, the old man and the old woman,
 helped *Modi and his family*.’ [RD, 2:111]

mene

mene is the most elusive of the classifiers, and in fact, its status as a classifier is questionable. In addition to occurring with quantifiers of different types, it can also apparently be used with any type of lexical head. The following presents a number of examples where it seems to function as a typical classifier:

199. *moŋ mene daḥom* ‘one sacrifice’
ghol mene seyan ‘ten village elders’
muḍu mene khaḥiya ‘one Kharia (person)’
200. *je[ʔ]=ga duʔkho muḍu mene=yaʔ=te agar*
 so=FOC misfortune one.HUM CLASS=GEN=OBL if
ḍe=na laʔ=ki laʔ baŋcay=na=?
 come=INF IPFV=MID.PST then save=INF=GEN
thom sub bhere sampere=ga aw=na laʔ=ki=kiyar.
 PURP all time ready=FOC QUAL=INF IPFV=MID.PST=DU
 ‘Thus if misfortune befell one of them (= came to *one’s [home]*),²²
 then the two remained ready at all times to save [each other].’
 [HJPa:47]

²² See 5.2 above on “double-case” marking.

Unlike other classifiers, *mene* is not restricted to use with numerals or even quantifiers. For example, it is found with demonstratives:

201. *la? atha jhelob, ikud=ga jhelob ro ho mene=ga cakar*
 then heavy long very=FOC long and that CLASS=FOC wide
arbo samudar=te=jo odo?=ki ho=ki “bapre!”
 Arab ocean=OBL=ADD other=PL that=PL wow!
gam=o?=ki.
 say=ACT.PST=PL
 ‘Then at *that* heavy, long, very long and so very wide Arabian Sea
 the others all said “Wow!” ’ [MT, 1:30]

It is even found following a genitive attribute (202) or an adverbial (203):

202. *ber=a[?] mene takat aw=ki?*
 who=GEN CLASS power QUAL=MID.PST
 ‘Whose [men] were powerful (= power was *whose*)?’ [MS, 1:202]
203. *khõde?j=ko doko=te=ga mukum=ki, tay ho bo?=te=ga*
 a.while=CNTR sit=OBL=FOC doze.off=MID.PST then that place=OBL=FOC
 “*khõre?j mene dhalanpe=ta=ijn*” – *gam=o? ro*
 a.while CLASS take.a.nap=MID.PRS=1SG say=ACT.PST and
gita?=kon leme?d god=ki.
 lie.down=SEQ sleep C:TEL=MID.PST
 ‘Sitting for a while, he began to doze off (= he dozed off), so he
 said [to himself] at that place “I’ll sleep *for a while*.” and he lay
 down and fell asleep.’ [BB, 2:27]

It is conceivable that *mene* derives from the plural marker *man* in Sadri, although the fact that it is also compatible with the numeral ‘one’ in Kharia (e.g. (199), (200)) and singular entities in general (e.g. (201)) would appear to argue against this. Consider, however the following example. Here *mene* does not follow the cardinal number *ubar* ‘two’ but rather the lexical head, *kuyu* ‘pot’, giving it the appearance of a plural marker:

204. ... *ubar kuyu mene pe?* ‘two pots of rice’ [HJPa:147, line h:5]
 two pot class rice

Finally, *mene* is also found in contexts which translate as ‘so much, as much as’, again with a genitive attribute:

205. *umbo? la? jahā bhēre konseldu?=ki=ya? mene=jo*
 otherwise INDEF time WOMAN=PL=GEN CLASS=ADD
umay bu?jhi bha?=ta.
 NEG.3PL understand SUD=MID.PRS
 ‘But otherwise, at any time, [men] do not understand as quickly as
 much as women do.’ [Kerkeṭtā, 1990: 12]

Whatever its etymology, as *mene* generally does follow a quantifier, it will tentatively be considered a classifier here. Its exact status awaits further study.

=o

=o is the general classifier and is used with all types of contentive morphemes, both human and non-human. It is by far the most common classifier.

206. *duy=o daru* ‘two trees’ *chaw=o soreṅ* ‘six stones’
tin=o bhai ‘three brothers’

=o and *j(h)an* (see above) can be combined with human reference, although this is seldom:

207. *pāc=o jan bhai* ‘five brothers’

However, =o is not compatible with other classifiers:

208. **pāc=o tho tebul* ‘five tables’

Perhaps this usage is related to the collective function of =o, described in the last section, although if so, it is unclear why its use is ungrammatical in combination with other classifiers.

Finally, note that units of time, such as 'minute', 'hour', etc., are not classifiers in Kharia, as they can combine with the classifier =*o*: *das=o minaŋ* 'ten minutes'

taŋ, thaŋ, thoŋ

This classifier is equal in function to *tho* (see below) and, according to speakers we consulted, is simply an alternative pronunciation of *tho*.

209. *thamke=na=pe se la... ..al=e=pe se la,*
quiet=MID.IRR=2PL please VOC stay=ACT.IRR=2PL please VOC

ubar thaŋ ghoŋa yo=ta=kiyar.

two CLASS horse see=MID.PRS=DU

'Everybody please be quiet, please stay put, two horses can be seen.'

[Kerkettā, 1990: 1]

210. *hin=a? thom ponomosor ubar taŋ maha*
that=GEN for God two CLASS big

patar=te bay=o?... .

light=OBL make=ACT.PST

'Therefore God made two great lights

[HPa:100]

Malhotra (1982: 67) and Biligiri (1965: 202) both write that this classifier is restricted to use with *ubar*, which fits well with this author's corpus data. However, in interviews, speakers also accepted its use with other cardinal numerals.

tho

Native speakers we consulted claimed that *tho*, which has apparently been borrowed from Sadri (cf. Jordan-Horstmann, 1969: 155), is used only with non-humans, whether animate or inanimate. Before *tho*, *chaw* is realized as *che*. Note that *tho* can also appear with quantifiers such as *kai* 'a few' which are not cardinal numbers:

211. *pāc tho ore?j* 'five oxen' *che tho soreŋ* 'six stones'
das tho tebul 'ten tables' *kai tho ompay* 'a few rivers'

However, examples in spoken texts were found in which *tho* is also used with reference to humans. In the following example, the southern speaker uses the human classifier *jhan* (see above) for his sons but *tho* for his

daughters, although later on in the same story (line 29) he also speaks of the nine sons as *naw tho bet=dom*:

212. *la[ʔ] hondo=waʔ naw jhan beta, naw tho beti*
 then Hondo=GEN nine CLASS son nine CLASS daughter
aw=ki=may.
 QUAL=MID.PST=3PL
 ‘Then Hondo had *nine* sons and *nine* daughters.’ [MS, 1:23]

Biligiri (1965: 203) notes that this classifier is restricted to numerals above ‘3’. However, the intuition of speakers we worked with was that *tho* is merely an alternative pronunciation of *taŋ / thaŋ / thoŋ*, which is compatible with *ubar* ‘2’ as well. This issue requires further study.

5.10 MODIFICATION

5.10.1 Morphologically simple modification (“adjectives”)

If one uses only morphosyntactic criteria as a guide to determining parts-of-speech, then there is little evidence for assuming a separate class of adjectives in Kharia, as argued in Chapter 4. Instead, from a purely structural point of view it is more convenient to consider simple attribution, i.e., attribution in which there is no (partially) finite TAM/PERSON-syntagma, to be expressed by the masdar, described in Section 6.6.2.1: Contentive morphemes appear in their basic form if they are polysyllabic and are reduplicated if monosyllabic. Consider the following forms:

213. <i>Example</i>	<i>Meaning (in middle voice)</i>	<i>Masdar</i>
<i>botoŋ lebu</i> ‘frightened person’	<i>botoŋ</i> ‘become afraid’	<i>botoŋ</i>
<i>ḍoko thãro</i> ‘settled place’	<i>ḍoko</i> ‘sit down; settle’	<i>ḍoko</i>
<i>konon bhai</i> ‘little brother’	<i>konon</i> ‘become small’	<i>konon</i>
<i>maha daru</i> ‘the big tree’	<i>maha</i> ‘become big’	<i>maha</i>
<i>ruʔ-ruʔ kaʔbto</i> ‘open door’	<i>ruʔ</i> ‘become open’	<i>ruʔ-ruʔ</i>
<i>rusuŋ oʔ</i> ‘red house’	<i>rusuŋ</i> ‘become red’	<i>rusuŋ</i>

Note that reduplication here is purely phonologically motivated. This is especially clear in cases such as *ḍoko ro ḍhoʔ-ḍhoʔ thāro* [settle.down and take-RDP place] 'the place where [they] settled down and which they took' As *ḍoko* is bisyllabic, it is not reduplicated whereas the monosyllabic *ḍhoʔ* is.

Although no longer in current use, there also appears to have been a contentive morpheme *saŋ* 'become yellow', which had the form *saŋ-saŋ* when used attributively, i.e. 'yellow' (in addition to 'turmeric'), and which is given with this meaning in Floor et al. (1934: 112).²³ This form has now been replaced by the Sadri term *piyar* 'yellow', as all color terms except *osel* 'white', *mogher* 'black' and *rusuŋ* 'red' have been replaced by the corresponding Sadri equivalents,²⁴ and in current usage *saŋ-saŋ* means only 'turmeric', a further meaning which it has apparently always had (cf. the data in Anderson & Zide, 2002: 58).

On the other hand, any contentive morpheme that can be used attributively can also be used to denote a change of state (middle voice) or a transitive telic action (active voice) and *vice versa*, hence there seems to be no need to assume an adjectival class in Kharia.

There are only few monosyllabic contentive morphemes which do not fit this pattern and which do not reduplicate,²⁵ almost all of which are loans from Indo-Aryan; loans from Indo-Aryan never reduplicate, with the exception of *col* 'go' Cf. forms such as *khus* '(become) happy', *lil* '(become) blue' and *bes* '(become) good', e.g. *bes dinu* 'good day' As the glosses indicate, it would also be incorrect to consider these contentive morphemes to be members of a closed adjectival class, as they can appear not only in attributive function but—like any other contentive morpheme—also in predicative function.

The repetition of attributes can denote both intensity ('very') and plurality. When used to denote plurality, the semantic head is not generally marked for number, as with *konon konon murti* in the following example:

²³ This form is given there as *sang* and is said to derive from *sarsang sarson* (?) "≡ which has a yellow flower. yellow." I am not familiar with either of these two forms, neither of which is listed in Floor et al. (1934), and also do not find them mentioned elsewhere in the literature.

²⁴ It is interesting to note here that the three terms which have not been replaced by the corresponding Sadri terms are the three left-most color terms in the two hierarchies of basic color terms found in Berlin & Kay (1969: 104) and the term for 'yellow', which has only recently been borrowed, immediately follows these three in one of their two hierarchies.

²⁵ But they do repeat to denote plurality, etc. See the following discussion in the main text.

214. *ho=kaṛ lo?kha ḍoṭh=o? ro lo?kha ḍo?ḍ=na lo?ḍho*
 that=SG.HUM soil take=ACT.PST and soil take=INF after
i=gḥay aḍi=ya? ruṭ buṇ=ga konon konon murti
 what=way ANAPH=GEN form INST=FOC small REP statue
baṇ=o?
 make=ACT.PST
 ‘He took soil and, after taking soil he made *small statues* in (= with)
 his [own] form.’ [AK, 3:10]

It is important here to note that this repetition differs from the reduplication found with the *masdars*: Reduplication is a purely grammatical process which serves to derive bisyllabic words from monosyllabic contentive morphemes in certain environments. That is, in this process phonological words are being created, not repeated. In the repetition described here, however, an element is being repeated which is already a phonological word.

5.10.1.1 *Comparatives, superlatives and equatives*

Although comparatives, superlatives and equatives are, properly speaking, a matter of syntax in Kharia, for reasons of exposition and reference these three constructions will be dealt with in this section.

Comparatives and superlatives

Comparatives and superlatives—in all functions, whether attributive, referential or predicative—are formed as follows. The standard is followed by an ablative postposition and then by the contentive morpheme denoting a scalar property, which appears in its base form. The two constructions are identical except for the fact that the standard in the superlative construction contains a quantifier denoting entirety, such as *ḥaṛi* or *sou?b* ‘all’. Also, there is a distinct preference for the ablative postposition *tay* in the comparative construction and *se* in the superlative construction, although both may be used in both constructions.

Comparatives

215. *la? aḍi=ya? tay konon bḥai, beḥma, ho=kaṛ*
 then ANAPH=GEN ABL small brother Beḥma that=SG.HUM
oṇḍor=o?
 hear=ACT.PST
 ‘Then *his_i younger* brother, Beḥma_i, he_i heard [him_i.]’ [AK, 1:30]

216. *aba=ḍom=a? gam=na=? motabik=ga hapta dinu=ya?*
 father=3POSS=GEN say=INF=GEN according.to=FOC week day=GEN
lo?ḍho raʔa aba=ḍom=te moŋ hajar rupaya tay=jo
 after Rata father=3POSS=OBL one thousand rupee ABL=ADD
jughay ter=o?
 much give=ACT.PST
 'As his father had said (= according to his father's saying), after a week Rata gave his father *even more than 1,000 rupees.*'
 [BB, 2:59]
217. *teuwa?jo modi=te oʔo?=ki=ya? tay jughay=ga dular*
 nevertheless Modi=OBL other=PL=GEN ABL much=FOC love
karay=na laʔ=ki=kiyar.
 do=INF IPFV=MID.PST=DU
 'Nevertheless, they loved Modi *more than the others.*' [RD, 2:16]

Alternatively, the standard can be left implicit, as in the following example. Here *oʔo?* 'increase; and; more' is used to denote the (implicit) comparison:

218. *u soub haleit yo=kon modi oʔo?=ga uʔḍ=na*
 this all condition see=SEQ Modi more=FOC drink=INF
maʔe=yo?
 begin=ACT.PST
 'Seeing this entire condition, Modi began to drink *even more.*'
 [RD, 2:107]

Superlatives

219. *simra souʔb se maha beʔṭ=ḍom souʔb tay siŋ ho maʔa*
 Simra all ABL big son=3POSS all ABL first that cave
boʔ=te col=ki ro soreŋ=a? moʔjhi tay ʔaʔ-m-uʔ=na
 place=OBL go=MID.PST and stone=GEN middle ABL water-m-drink=INF
laʔ=ki.
 IPFV=MID.PST

‘Simra, *the eldest* brother, went to that cave *first of all* [*the brothers*] and he was drinking water from the middle of the stones.’

[AK, 1:71]

220. sou^ʔb se maha=ya^ʔ biha hoy=si^ʔ.

all ABL big=GEN wedding become=PERF

‘The oldest has married (= The marriage *of the oldest* has become).’

[AK, 4:9]

Note that these constructions are used with all scalar contentive morphemes and are not limited to contentive morphemes in attributive function. Thus comparatives and superlatives cannot be considered typical properties of “adjectives”, but merely of contentive morphemes denoting scalar concepts.

221. sou^ʔb se suru=te ponmesor ap aw=ki.

all ABL beginning=OBL God Father QUAL=MID.PST

‘*In the very beginning* there was God the Father.’

[AK, 3:4]

Equatives

Equatives are formed by means of the postposition *lekhe* ‘like’, a loan word from Sadri.

222. ho=wa^ʔ lekhe su^ʔkho t̪hãro atu=ga ro^ʔ kuy=e=nan

that=GEN like happy place where=FOC FOC find=ACT.IRR=1DU.INCL

re^ʔ

voc

‘Where will we ever find a place *as happy as that*?’

[AT, 3:2]

223. kinir d̪am=kon ra^ʔta thorek to^ʔ leri=ki, ro

forest arrive=SEQ Rata little time rest=MID.PST and

aba=d̪om=a^ʔ lekhe=ga jal ropa=kon dana chi^ʔta=yo^ʔ

father=3POSS=GEN like=FOC net spread.out=SEQ grain scatter=ACT.PST

ro daru sumbo^ʔ=te le^ʔd̪ go^ʔd̪=ki.

and tree base=OBL hide C:TEL=MID.PST

‘After arriving at the forest, Rata rested a little while, and *just like his father* he spread out the net, scattered grain [on it] and hid at the tree trunk.’

[BB, 2:25]

224. samudar lekhe *ompay=ki boṛe boṛe. la? ho=te*
 ocean like river=PL big REP then that=OBL(= 'there')
paro=na muskil heke.
 cross=INF difficult QUAL
 'The rivers are very full (= big), *like an ocean*. Then it is difficult
 to cross there.' [MS, 1:157]

5.10.2 Clausal adjuncts ("adverbials")

In principle there is no need to discuss "adverbials" separately in Kharia as they do not differ from other types of Case-syntagmas. This section therefore concentrates on a strategy which is quite often found in adjuncts and which would seem, at least at first sight, to be a kind of derivational adverbial marker.

One of the most common means of modifying a predicate or a clause is by adding *-bo?*, *-ḍa?* or *-son* to a contentive morpheme. These do not differ in meaning and the choice of one over the other is lexically determined. They can also occasionally combine, as in *lere?-ḍa?-son*. A few of the more common forms are presented here as well as two typical examples of their use:

- | | |
|-----------------------------|--|
| 225. <i>aṅjor</i> 'front' | <i>aṅjor-ḍa?</i> 'forwards' |
| <i>bes</i> 'good' | <i>bes-bo?</i> , <i>bes-ḍa?</i> 'well' |
| <i>dular</i> 'love' | <i>dular-bo?</i> 'lovingly' |
| <i>dhirom</i> 'become slow' | <i>dhirom-bo?</i> 'slowly' |
| <i>lere?</i> 'rejoice; joy' | <i>lere?-son</i> , <i>lere?-ḍa?-son</i> 'joyously' |
226. *ro tama am=pe u nāw kuṭum=te=ga sadi biha*
 and now 2=2PL this nine family=OBL=FOC marry marry
kerson=na=pe ro baru-bo? aw=na=pe.
 marry=MID.IRR=2PL and good-INTENS live=MID.IRR=2PL
 'And you will marry in only these nine families and you will live
 (very) well.' [AK, 1:69]
227. *hardinagar tay mu?=kon buli=ga buliga aṛmaṛay=ga*
 Hardinagar ABL emerge=SEQ wander=FOC REP go.back.and.forth=FOC
si?ḍ=ga col kan=ki=may aṅjor-ḍa?
 get.lost=FOC go CONT=MID.PST=3PL front-INTENS
 'After leaving Hardinagar, they continued on (= kept going *straight forward*), wandering, going back and forth and getting lost.'
 [MT, 1:117]

As the following example shows, however, this construction is not exclusively used to mark adverbials:

228. *maha kimin=ya?* *koleʔj-bo?* *yo=kon* *etwa konon*
 big daughter.in.law=GEN fight-INTENS see=SEQ Etwa small
beʔ=ɔom=a? *biha=ya?* *tiha* *karay=o?*
 son=3POSS=GEN marriage=GEN provision do=ACT.PST
 ‘Seeing how (*very*) *quarrelsome* the elder daughter-in-law was,
 Etwa arranged for the marriage of his younger son.’ [RD, 2:75]
229. *ho=kaʔ* *ikuʔd* *juhay* *lereʔ-son=ki.*
 that=SG.HUM very much joy-INTENS=MID.PST
 ‘S/he became *very happy*.’

What these apparent “adverbials” in fact express is a slight degree of “intensity”, that is, they denote that an event or state is somewhat “more intense” than the unmarked forms, similar to the function of *very* in English. Hence they are not restricted to adverbial function, although they are certainly compatible with this usage.

This semantic contribution is, however, very weak—often almost non-existent—so that it may well be that these markers are in the process of developing into true adjunct markers, although all speakers consulted agreed that they have a slightly intensifying meaning.²⁶

Other types of “adverbials”, such as temporal and locative adverbials, are typically expressed through attributive finite clauses with a locative or temporal head (230) or simply by a Case-syntagma in the oblique case (231).

230. *ho=kaʔ* *ʔhik* *biha=na* *laʔ=ki=may* *bhere*
 that=SG.HUM exactly marry=INF IPFV=MID.PST=3PL time
ɔam=ki.
 arrive=MID.PST
 ‘[He] arrived just as [they] were getting married (i.e., **at the they-were-just-getting-married time**).’ [HJPa:41]

²⁶ This would seem to have a direct correlate in several East and Southeast Asian languages. Confer the discussion of the modification of sentence adverbs in Chinese and Vietnamese in Bisang (1998: 692–4) with morphemes denoting ‘very’

231. *kahami lebu=ki khoʔi=ki=te kayom=ta=ki*
 story person=PL village.section=PL=OBL speak=MID.PRS=PL
 ‘The people tell [this] story *in the villages*.’ [AK, 1:2]

For further examples of “adverbial” modification, see the sections on “relative clauses” (7.6) and imperfective converbs and depictives (Sections 6.6.2.4 and 7.5.1.6).

THE TAM/PERSON-SYNTAGMA ('VERBS')

6.1 GENERAL INTRODUCTION

The Kharia predicate can conveniently be divided into two major structural types. The first consists of predicates of the type which Bloomfield (1933[1984]: 173) refers to as “narrative predication”, in which, e.g., the subject performs an action. In these predicates in Kharia, tense, aspect and mood are expressed through portmanteau morphemes which also simultaneously express basic voice (active vs. middle). Following this are the markers for person, number and honorific status. This type also includes three minor sub-types: The present perfect, the optative and the Past II, which do not mark for basic voice but are otherwise structurally identical. With the exception of the optative, all of these markers are enclitic, whereas the optative marker is a phonological word and can be separated from other elements, e.g., by the enclitic focal particles.

The second type of predication consists of two sub-types, both of which are subsumed under “equational predication” by Bloomfield (1933[1984]: 173). In the first of these two sub-types, which will be referred to here as “non-inherent qualitative predication”, the position of an entity is given or a statement is made about a temporary or non-essential characteristic of this entity. The second sub-type is what Maas (2004) refers to as “qualitative” predication, in which an inherent property of an entity is depicted (e.g. *I am a man, That is a tree*). “Inherent” should be taken here as meaning something akin to “central or most important characteristic” and will undoubtedly be subject at least to some extent to a particular speaker’s judgement. This type of predicate will be referred to here as “inherent qualitative predication” Both of these sub-types are expressed by markers for qualitative predication which, except in the present tense, have the same structure as narrative predicates, i.e., they are marked for TAM/basic voice and PERS/NUM/HON. In the present tense, however, they do not mark for basic voice. The qualitative predicative marker may also not be present, or “lacking”, resulting in what resembles a “nominal sentence” but with subtle semantic differences to examples in which an element translating as ‘be’ or ‘become’ is present.

The present section concentrates on the structure of the narrative predicate, with the various predicating categories presented in the traditional manner. Qualitative predication will be dealt with in Section 7.3, where evidence against considering the qualitative predicative markers to be copular forms will also be presented.

Narrative predication: An overview

The (simplified) maximal structure of the non-periphrastic, fully finite narrative predicate in Kharia, henceforth the “TAM/PERSON-syntagma”, due to its structure, is shown in Table 6.1, where “X” is the (possibly complex) semantic base. The causative may appear either as a prefix or as an infix, or both in double causatives. Note that with a TAM/PERSON-syntagma in the present perfect, there is no TAM/basic voice marker following the PERFECT marker; otherwise the structure is identical.

Table 6.1: The structure of the TAM/PERSON-syntagma
REC CAUS-X-<CAUS> v2=PERF=TAM/VOICE=PERS/NUM/HON

Table 6.2 presents the structure of the remaining two minor types of TAM/PERSON-syntagmas, which contain either the optative marker *gudu?* / *guru?* or the Past II marker =*kho?* but which are not marked for basic voice.

Table 6.2: The structure of the TAM/PERSON-syntagma in the optative /
Past II

REC CAUS-X-<CAUS> v2=OPT/PST.II=PERS/NUM/HON

“X” in Tables 6.1 and 6.2 can have the same maximal structure as that given in Table 5.2 in Chapter 5 (without case marking!), repeated here for convenience as Table 6.3:

Table 6.3: A schematic overview of the maximal common semantic base of
TAM/PERSON- and Case-syntagmas

GENITIVE DETERMINER DEM QUANT CLASS GENITIVE DETERMINER
Lexeme(s)=POSS=NUM/HON

The following presents once again a few illustrative examples from Sections 4.4 and 4.4.1 for ease of reference which show that the semantic base of the event-narrating predicate may be complex:

1. a. ubar rocho³b=ki=j_L
 two side=MID.PST=1SG
 'I moved to both sides (i.e., this way and then that).'
2. a. ho rocho³b=ki=j_L
 that side=MID.PST=1SG
 'I moved to that side.'
3. *bharat=ya?* *lebu=ki* bides=a? *lebu=ki=ya?*
 India=GEN person=PL abroad=GEN person=PL=GEN
 rupraj=ki=may.
 appearance=MID.PST=3PL
 'The Indians *took on the appearance of foreigners* (e.g. by living abroad so long).'
4. *māṅta tamaku ragday=ta²j*, *bādhna kaṭi²j saṅrakhi lekhe*
 Mangta tobacco rub=MID.PROG Bandhna little stubborn like
 han=ti²j u=ti²j khor=ta, *kaṭi²j jorsāy bul=si?*
 that=side this=side ITER=MID.PRS little strongly become.drunk=PERF
 'Mangta is rubbing tobacco, Bandhna is *walking here and there* as if he is being a little stubborn, and is quite strongly drunk.'
- [Kerkeṭṭā, 1990: 24]

Expanding the “v2” slot in Tables 6.1 and 6.2 for the most commonly found “v2” categories and assuming for ease of presentation that the semantic base of the TAM/PERSON-syntagma is a single contentive morpheme (here for ease of reference: “**Lexeme**”), we arrive at Table 6.4:

Table 6.4: A schematic overview of the maximal structure of the non-negated, non-periphrastic, morphologically fully finite TAM/PERSON-syntagma in Kharia (simplified)

REC CAUS-**Lexeme**-<CAUS> EXCESS BEN CONT/ITER/S:ITER PASS/REFL AUTOPOES
 TEL=PERF=TAM/VOICE=PERS/NUM/HON

Only the causative marker is an affix, while the markers to the right, beginning with the perfect, are all enclitic (cf. Chapters 2 and 3). These enclitic markers DEFINE the TAM/PERSON-syntagma as such and attach directly to the rightmost element of the semantic base, as shown in examples (1)–(4),

and may not appear together with case marking (other than the genitive) or with adpositions.

6.2 PERSON / NUMBER / HONORIFIC MARKING

Kharia is a nominative/accusative type language in terms of personal marking on the TAM/PERSON-syntagma. PERS/NUM/HON-marking on the TAM/PERSON-syntagma always refers to the “subject”, which is either S, the “intransitive subject”, or A, “transitive subject” or rather, the highest entity, semantically speaking, in a transitive clause.¹ Table 6.5 gives an overview of subject-markers on the TAM/PERSON-syntagma.

As we shall see in Section 7.1, where the notion of subject is discussed in greater detail, this does not necessarily mean that the TAM/PERSON-syntagma AGREES in terms of person, number and honorific status with an explicitly mentioned “subject”: Subject marking in Kharia is what is often referred to as “semantically” based—the personal marking on the TAM/PERSON-syntagma itself is best considered the subject of the clause, while the explicitly mentioned “subject” is in apposition to this element.

Table 6.5: Person/Number/Honorific Marking on the TAM/PERSON-syntagma

	<i>Singular</i>	<i>Dual (/HON)</i>		<i>Plural</i>	
		<i>Inclusive</i>	<i>Exclusive</i>	<i>Inclusive</i>	<i>Exclusive</i>
1	= <i>(i)n</i> / = <i>(i)ŋ</i>	= <i>naŋ</i>	= <i>jar</i>	= <i>niŋ</i>	= <i>le</i>
2	= <i>(e)m</i>		= <i>bar</i>	= <i>pe</i>	
3	–		= <i>kiyar</i>	= <i>ki</i> / = <i>may</i>	

¹ For a discussion of A, S and O, cf. Dixon (1994: 6ff.). For our purposes, the following definition from Andrews (1985: 68) will suffice, substituting “Case-syntagma” for “NP” and “TAM/PERSON-syntagma” for “verb”. “PTV” here is an abbreviation for “primary transitive verbs”, the class of two-argument verbs which take an agent and patient:

If an NP is serving as argument of a two-argument verb, and receiving the morphological and syntactic treatment normally accorded to an Agent of a PTV, we shall say that it has the grammatical function A; if it is an argument of a verb with two or more arguments receiving the treatment normally accorded to the Patient of a PTV, we shall say that it has the grammatical function O. A sentence is called ‘transitive’ if it has A and O functions in its syntactic structure, ‘intransitive’ if one or both functions is missing. [...] An NP in an intransitive sentence that is receiving the treatment normally accorded to the single argument of a one-argument predicate will be said to have s-function [...].”

As (5) and (6) show, the TAM/PERSON-syntagma usually seems to agree with the apparent “subject”, which is either S or A, in terms of person, number and honorific status:

5. *la?* *sou?b=ga* *bhai=ki* *ho* *khajar* *tar=na* *ghađ* *juda*
 then all=FOC brother=PL that deer kill=INF PURP separate
 S

juda *mu?=ki=may,* *kinir=te.*
 REP emerge=MID.PST=3PL forest=OBL

Predicate

‘Then all the brothers set out separately to kill that deer, into the forest.’ [AK, 1:15]

6. *muda.* *ha?đo=ki* *khariya* *so?j=te=may,* *adha=ki*
 but half=PL Kharia understand=ACT.PRS=3PL half=PL
 A O Predicate₁ A

khariya *umay* *so?j=te.*
 Kharia NEG.3PL understand=ACT.PRS

O Predicate₂

‘But. half understand Kharia, half don’t understand Kharia.’

[AK, 5:6]

The 3rd person plural marker =*may* is generally found after the middle past marker =*ki*, thus avoiding the sequence =*ki=ki* ‘=MID.PST=3PL’, although this sequence is occasionally found. =*ki* is generally used elsewhere although it is often replaced by =*may* in these environments as well, especially after the perfect marker =*si?*. Thus =*may* and =*ki* are identical in meaning when they signal a 3rd person plural subject: The choice of one over the other in these environments is largely speaker-specific. Cf. e.g. the following example, where the two appear side-by-side:

7. *tay* *lam=na* *lamna* *lamna* *ho=ki* *ho* *jinis=te,* *khajar,*
 then search=INF REP REP that=PL that animal=OBL deer

kuy=o?=ki *ro* *tar=o?=may.*
 find=ACT.PST=PL and kill=ACT.PST=3PL

‘Then searching, searching, searching they *found* that animal, a deer, and *killed* [it].’ [AK, 1:53]

The only environment in which they are not in free distribution is with the negative morpheme *um*: When a clause with a 3rd person plural subject is negated, only *umay* (< **um=may* ‘NEG=3PL’) may be used, never **um=ki* (6.7.1).

Note that, with the exception of *=may* ‘=3PL’, the agreement markers of the 3rd persons are, properly speaking, number markers and are identical with number marking found within the semantic base itself. E.g. *lebu*(=Ø) ‘man, person’, *lebu=kiyar* ‘men, people’ (DU) and *lebu=ki* ‘men, people’ (PL) (5.4). Hence these markers on the TAM/PERSON-syntagma are glossed here as singular (SG), DU(al) and PL(ural), respectively, but not for person. Only *=may* will be glossed for person, i.e., ‘3PL’

6.3 TYPES OF SEMANTIC BASES IN THE TAM/PERSON-SYNTAGMA

There are many types of semantic bases found in Kharia, ranging from simple contentive morphemes—possibly with alternating forms for middle and active voice and possibly marked for reciprocity or the causative—to more complex predicate bases consisting of two or more different semantic sub-bases, in addition to other types. These will now be discussed in more detail.

6.3.1 *Single contentive morpheme as the semantic base*

Simple, non-derived semantic bases

This is the simplest and by far the most common type of semantic base of the TAM/PERSON-syntagma in Kharia. In non-periphrastic forms, the markers shown in Tables 6.1 and 6.2 above directly follow the simple contentive morpheme:

8. *gam=te=ki* ‘they say’

say=ACT.PRS=PL

9. *ɖaŋ goʔh=oʔ* ‘he sent [them] off’

send C:TEL=ACT.PST

10. *muʔ goʔd=ki=may* ‘they went out’

emerge C:TEL=MID.PST=3PL

Masdar as the semantic base

The semantic base of the TAM/PERSON-syntagma may also be a masdar (6.6.2.1). In this construction, monosyllabic forms reduplicate whereas polysyllabic forms do not reduplicate.

This construction has a number of different meanings, including remote past, habitual, or that an event took longer than usual to occur. Regardless of whether the underlying (underived) contentive morpheme otherwise appears only in the middle, the active, or whether it can appear in both basic voices in its unmarked use, in this construction all TAM/PERSON-syntagmas can only appear with markers of the middle voice (cf. the discussion of the generic middle in Section 6.4.2.2).

11. *lam=o?*

search=ACT.PST

'S/he looked for [it].'

*Underived morpheme as semantic base*12. *lam-lam=ki.*

search=MID.PST

'S/he looked for [it] for a long time.'

*Masdar as semantic base*13. *gore²j hakre=yo?*

ox grunt=ACT.PST

'The ox grunted.' (unmarked)

*Underived morpheme as semantic base*14. *gore²j hakre=ki.*

ox grunt=MID.PST

'The ox grunted.' (a long time back)

Masdar as semantic base

The masdar also often serves as the semantic base in TAM/PERSON-syntagmas marked by one of the "durative" markers (6.5.2).

6.3.2 *TAM/PERSON-syntagmas with multiple semantic sub-bases*

In these TAM/PERSON-syntagmas, the complex semantic base consists of one or more juxtaposed semantic sub-bases, each of which may be either a simple, underived contentive morpheme or a derived form, consisting of an underlying morpheme and derivational marking such as the causative, reciprocal, etc. Grounding markers such as PERSON/NUMBER/HONORIFIC STATUS and TAM/VOICE are only marked once and have scope over the entire

phrase. It is important to note that complex TAM/PERSON-syntagmas have multiple *semantic sub-bases*, not “roots” or even (underived) contentive morphemes: E.g., causative markers are marked on each semantic sub-base to which they apply. Although less common than TAM/PERSON-syntagmas having one single semantic base, complex bases of this type are nevertheless quite common in Kharia. Semantically speaking, these semantic bases can broadly be divided into two general types:

- In the first type, the two or more semantic sub-bases have simultaneous reference, often referring to the same event. Note that (15) is one of the very seldom cases of a TAM/PERSON-syntagma with a complex semantic base consisting of three contentive morphemes.

15. *ro tama am=pe u nãw kuṭum=te=ga sadi biha*
 and now 2=2PL this nine family=OBL=FOC marry marry
kersonj=na=pe ro bes-boʔ, baru-boʔ aw=na=pe.
 marry=MID.IRR=2PL and good-INTENS good-INTENS QUAL=MID.IRR=2PL
 ‘And you *will marry* in only these nine families and you will be very happy.’ [AK, 1:69]

16. *am=pe khaṛiya hoy=kon u=jeʔ yo sohe=ye=peʔ*
 2=2PL Kharia become=SEQ this=SG.NHUM see suffer=ACT.IRR=2PL
 ‘You, as (= having become) Kharia, [*can*] you bear to see this? (= will you see [*and*] suffer)’ [Kerkeṭṭā, 1990: 5]

As the following example shows, semantic bases of this type are not restricted to finite forms:

17. *u=kāṭ=aʔ thoŋ buṛha=ki=te ikuʔḍ ikuʔḍ dhain*
 this=SG.HUM=GEN for old.man=PL=OBL very REP thank
mane=na ayiʔj.
 honor=INF QUAL.PRS
 ‘For this the elders should be thanked and honored (= it is to *thank* [*and*] honour the elders for this).’ [MT, 1:183]

Occasionally, the first morpheme / sub-base primarily serves to specify the manner in which the action was carried out:

18. *banda bunui=a? moŋ caŋuwa go?*
 castrated pig=GEN one piece.of.shoulder carry.on.shoulder
 ol=e=ki...
 bring=ACT.IRR=PL
 'They will bring by carrying on their shoulders the shoulder of a castrated boar' [HJPa:153, b:4]

19. *og=a? kole? leŋ col kan=ki.*
 house=GEN parrot fly go CONT=Mid.PST
 'The parrot of the house has flown away.' [HJPa:219, 25]

Occasionally, the co-occurrence of the two sub-bases / morphemes has a more-or-less lexicalized meaning:

20. *yohan ho=ki ti?j gam o-y-eŋ=o? no:*
 John that=PL side say CAUS-y-return=ACT.PST Cmpl
 'John answered them, saying (= say-returned)' [HJPa:88, ln. 26]

– In the second, more common type, the sub-bases refer to different, usually sequentially occurring events. The first sub-base / morpheme then refers to the event which occurred first:

21. *ciŋtay ol=e=m!* 'Pinch (him/her) and bring (him/her)
 pinch bring=ACT.IRR=2SG here!'
22. *am=pe leŋ-leŋ jãut tar ol=e=pe.*
 2=2PL fly-RDP animal kill bring=ACT.IRR=2PL
 'You kill [and] bring [back] birds (= flying animals).' [MT, 1:131]
23. *tay ađi=te=jo batay=o? no "dhãy[=e] am, han*
 then ANAPH=OBL=ADD tell=ACT.PST Cmpl hurry=ACT.IRR 2SG that
maŋa bo?=te da[?] [a]yi?j. uđ de=na=m."
 cave place=OBL water QUAL.PRS drink come=Mid.IRR=2SG
 'Then he said to him, too, "Hurry, at the cave there is water. Drink and come [back]."' [AK, 1:38]

24. *agar am=a? bhabru onḍor=kon kiṛo? de=na la?*
 if 2SG=GEN bark hear=SEQ tiger come=MID.IRR then
boriya=te=ga tar jlog=e.
 both=OBL=FOC kill eat=ACT.IRR
 ‘If, having heard your barking, the tiger comes, he will *kill and eat* [us] both.’ [BB, 3:40]
25. *muda je kamu=na la?=ki ho=je ? u²ḍ*
 but CREL eat=INF IPFV=MID.PST that=SG.NHUM drink
jo?²=kon=ga pal=na la?²=ki.
 eat=SEQ=FOC finish=INF IPFV=MID.PST
 ‘But whatever he earned, it was just enough to live on (= he would finish, *having eaten and drunken*).’ [RD, 2:36]

We occasionally find a resultative interpretation in such TAM/PERSON-syntagmas, in which the second sub-base / morpheme refers to the resultant state:

26. *ma=ḍom kunḍu? kosu buṅ ṛyam patar=na*
 mother=3POSS child sickness INST cry light=INF
la?²=ki=kiyar.
 IPFV=MID.PST=DU
 ‘Mother and child *were becoming pale through crying* due to the sickness.’ [RD, 2:106]

As the next example (second and third lines) shows, the scope of the causative is restricted to the sub-base it marks:

27. *musniṅ jhaṛi kōḍpuru?²=ki jume kati²b=kon socay=o?²=ki*
 one.day all man=PL assemble gather=SEQ think=ACT.PST=PL
no “ho dano=te i=gud²=ga tar
 CMPL that demon=OBL what=like=FOC kill
o?²-gur=e=niṅ?”
 CAUS-fall=ACT.IRR=1 PL.INCL
 ‘One day, all the men *gathered* and thought “How will we kill (= *kill and cause to fall*) that demon?”’ [MT, 1:66]

The juxtaposition of sub-bases to form the overall semantic base of TAM/PERSON-syntagmas would appear to be restricted to cases in which the two or more events are considered in some way directly connected to one another, similar to the use of the sequential converb or “conjunctive participle” (6.6.2.3) in many South Asian languages.² It is regularly found with contentive morphemes which refer to actions which generally occur together (*no? uq* ‘eat drink’) but may apparently also be used for any combination of events which the speaker wishes to convey as related. Further research on this topic is necessary.

6.3.3 Other types of semantic bases

A number of other types of semantic bases differ somewhat from the types described in the last sections. These will now be discussed individually.

ECHO-words

The semantic base of a TAM/PERSON-syntagma may also be formed by one of the constructions discussed in Section 4.7 under the heading “echo-words” As described there, a number of the “echo-words” would seem to have developed from the juxtaposition of what were once contentive morphemes, similar to the type described in the last section, but in which the second morpheme has since lost its lexical meaning.

28. *o?=te* *ap=d̥om* *teinko* *arkhi* *un=sig=e*
house=OBL father=3POSS a.little liquor place=PERF=ACT.IRR
- ho=je?=ko* *ho=kaɾ* *capu* *karba¹d̥=kon* *u¹d̥*
that=SG.NHUM=CNTR that=SG.HUM rummage ECHO=SEQ drink
- ida?¹=na* *la?¹=ki*
ECHO=INF IPFV=MID.PST

‘If his father has placed some liquor in the house, then he would rummage through everything [until he found it] and then drink it all up.’ [RD, 2:25]

² Cf. Peterson (2002b: 115ff.) for a brief discussion of this topic.

Predicates containing buŋ ‘INST’

This type of semantic base is generally found in predicates denoting psychological states. The following two examples illustrate the basic structure:

29. *gupa lebu duʔkho buŋ ikuʔd juŋhay goʔd=ki.*
 guard person sorrow INST very much C:TEL=MID.PST
 ‘The shepherd *became very depressed.*’ [RD, 1:18]

30. *ho simkom=te yo=kon ho=ki lereʔ buŋ i=ghay*
 that star=OBL see=SEQ that=PL joy INST what=way
goʔd=ki=may.
 C:TEL=MID.PST=3PL
 ‘Upon seeing that star they *rejoiced greatly.*’
 [Kullū, 1992: 3 = Matthew, 2:10]³

The two semantic bases have the “underlying” form, i.e., the form they would have were they not the semantic base of a TAM/PERSON-syntagma, *ikuʔd juŋhay duʔkho buŋ* ‘with much sorrow’ and *i=ghay lereʔ buŋ* ‘with such (= what=way) joy’, respectively.

Recall from Section 4.4 that predicative marking may attach to any semantic base unless it is already marked for “case”, i.e., the oblique marker =*te* or a postposition. In the case of the two “underlying” semantic bases above, however, we find *buŋ*, a postposition, as the last element—as such, predicative marking may not simply be added to these units.

Intuitively at least, this construction can be understood as a simple “re-ordering” of the constituents of the underlying semantic base: By moving the lexical head in these examples along with the postposition to the front, the semantic base no longer ends in a “case” marker.

Although this explanation may be intuitively appealing, a more in-depth study would undoubtedly shed more light on this construction. For further information on the status of this construction, especially the problems which arise by assuming a “hidden verb”, cf. Section 4.4.

³ Translation in the Saint Joseph Edition of *The New American Bible*: “They were overjoyed at seeing the star, . . .” (Catholic Biblical Association of America, 1970: The New Testament, 6).

Alternating forms

This group consists of a large number of contentive morphemes which all appear to have been borrowed from Sadri and which have two different forms, one in *-e* for the middle voice and one in *-ay* ([aɛ]), less often *-a*, for the active, with a few minor irregularities, such as *badle* / *badli* and *buʔjhi* / *bujhay*. Table 6.6 presents a few examples.

Table 6.6: The *-e/-ay* alternation

<i>Middle form</i>		<i>Active form</i>	
<i>bagɽe</i>	'be(come) bad'	<i>bagɽay</i>	'spoil (TR), ruin'
<i>badle</i>	'change' (ITR)	<i>badli</i>	'change' (TR)
<i>buʔjhi</i>	'understand'	<i>bujhay</i>	'explain'
<i>cale</i>	'function; go'	<i>calay</i>	'drive (a car, etc.)'
<i>chuʔɽe</i>	'leave' (ITR)	<i>chuʔɽay</i>	'leave' (TR)
<i>khole</i>	'open' (ITR)	<i>kholay</i>	'open' (TR)
<i>salge</i>	'burn, catch fire'	<i>salgay</i>	'light a fire'

The *-e* (written <y> after <a>) in Kharia in both forms derives from the infinitive marker in Sadri. Although this form in modern Sadri is *-ek*, it is virtually certain that *-ek* derives from an earlier *-e* plus the genitive marker =*k*, functioning as a nominalizer (cf. Peterson, forthcoming). It is this earlier form which has been borrowed into Kharia. This marker now serves to mark the contentive morpheme as a borrowed morpheme.

The status of *-a* in Kharia is more problematic. In Sadri, the suffix *-a* is, among other things, a causative marker, although its status there is still somewhat unclear.⁴ In Kharia, *-a* is not productive outside of these contentive morphemes.

While all of these contentive morphemes appear to be borrowings from Sadri, many of the forms marked for *-ay* in Kharia in fact derive from verbs in Sadri which are not marked there by *-a* with the same meaning. Cf. Kharia *karay* 'do' but Sadri *kar-* 'do', and Kharia *socay* 'think' but Sadri *soc-* 'think'.⁵

⁴ According to Jordan-Horstmann (1969: 56f), *-a* is a causative marker, in addition to its functions as a verbalizer, causative passive, and reflexive marker. According to Nowrangi (1956: 115), the same suffix is also a passive marker. Cf. (adapted from Nowrangi, 1956: 114f.) *sun-* 'hear', *sun-a-* 'be heard', *piy-* 'drink', *piy-a-* 'cause to drink' and *nac-* 'dance', *nac-a-* 'be made to dance'

⁵ This is not true of all morphemes here, though. Cf. Kharia *coray* 'steal' from Sadri *cor-a-* 'steal', which derives from the noun *cor* 'thief' (cf. Jordan-Horstmann, 1969: 56).

It would be incorrect to assume that *-a* is simply a causative marker or transitivity marker in Kharia: *-a* is also found in a number of contentive morphemes which can appear only in the middle or active, but not in both, such as *khisay* ‘become angry’ (middle only) or *chikay* ‘sneeze’ (active only). These forms do not have a counterpart marked in *-e*, and neither fits the usual description of a “causative verb”

While the exact status of *-a* in Kharia is difficult to determine, if in fact there is one, it is tendentially restricted to contentive morphemes denoting events whose subjects are normally animate, whereas predicating contentive morphemes without the *-a* (i.e., those in *-e*) generally have inanimate subjects. Table 6.7 presents a number of non-alternating contentive morphemes in *-ay* from our data. This table clearly shows that at least most of these predicating contentive morphemes require an animate subject.

Table 6.7: Non-alternating contentive morphemes in *-a(y)*

<i>batay</i> ‘tell’	Active	<i>jatay</i> ‘press’	Active
<i>belay</i> ‘write’	Active	<i>karay</i> ‘do’	Active
<i>biṛay</i> ‘test’	Active	<i>kuday</i> ‘chase’	Active
<i>boṭha</i> ‘soak’ (TR)	Middle	<i>khisay</i> ‘become angry’	Middle
<i>bhunbhunay</i> ‘buzz (of bees)’	Middle	<i>lagay</i> ‘plant seeds’; ‘throw color (on Holi)’	Active
<i>bhuriyay</i> ‘become still’	Middle	<i>maṛay</i> ‘put, place’ ⁶	Active
<i>cerberay</i> ‘chirp’	Middle	<i>melay</i> ‘leave’ (ITR/TR)	Active
<i>coray</i> ‘steal’	Active	<i>mesay</i> ‘mix’ (TR)	Active
<i>chekay</i> ‘stop’ (TR)	Active	<i>paday</i> ‘fart’	Active
<i>chikay</i> ‘sneeze’	Active	<i>socay</i> ‘think’	Active
<i>dhukay</i> ‘blow, fan’	Active ⁷	<i>ṭangay</i> ‘hang up’	Active
<i>dhumkay</i> ‘threaten’	Active	<i>ṭhakay</i> ‘deceive’	Active
<i>ghabṛay</i> ‘become confused’	Middle	<i>ṭhisay(y)</i> ‘call loudly’	Active
<i>ghinay</i> ‘hate’	Middle	<i>ṭharṭharay</i> ‘tremble (of people)’	Middle
<i>hagay</i> ‘defecate’	Active	<i>uṛay</i> ‘waste’	Active

This hypothesis is indirectly confirmed by those contentive morphemes ending in *-ay* which have been borrowed from Indo-Aryan which can occur in both the middle and the active, but which have only one form (in *-ay*) in both classes. Crucially, both middle and active forms require

⁶ Although there is a morpheme with the form *māre* and the meaning ‘come to a stop’, I do not believe that these two are related to one another synchronically, although perhaps diachronically.

⁷ Also possible in the middle voice, with no apparent semantic difference, but speakers preferred the active.

an animate subject. Examples of these contentive morphemes are the presented in Table 6.8. Note that a few, however, such as *gajgajay* 'overflow', cannot be accounted for by this analysis.

Table 6.8: Contentive morphemes in *-ay* in both Middle and Active

<i>Contentive morpheme</i>	<i>Middle</i>	<i>Active</i>	
<i>akbakay</i>	'be(come) nervous'	'make someone nervous'	
<i>cilay</i>	'shout'	'shout'	<i>No apparent semantic difference</i>
<i>cunay</i>	'be(come) selected'	'select'	
<i>danday</i> ⁸	'become amazed'	'amaze'	
<i>gargaray</i>	'clear of the throat (unintentional)'	'gargle; clear the throat (intentional)'	
<i>gohray</i>	'ask forgiveness'	'ask forgiveness'	<i>No apparent semantic difference</i>
<i>khuray</i>	'shave' (ITR)	'shave' (TR)	
<i>manjay</i>	'clean oneself'	'clean' (TR)	
<i>manay</i>	'be flattered'	'flatter'	
But: <i>gajgajay</i>	'overflow'	'overfill'	

Finally, this analysis is further supported by those morphemes which have been borrowed from Sadri and which have only one form in both voices but, unlike the examples in Table 6.8, these morphemes have only the form in *-e* and do not require an animate subject. Table 6.9 presents a few examples.

Table 6.9: Contentive morphemes in *-e* in both Middle and Active

<i>Contentive morpheme</i>	<i>Middle</i>	<i>Active</i>	
<i>māre</i>	'come to a stop'	'come to a stop'	<i>No apparent semantic difference</i>
<i>pase</i>	'come close'	'bring close'	
<i>tike</i>	'last a long time'	'last a long time'	<i>No apparent semantic difference</i>

⁸ The form *dande* is possible but speakers preferred the use of *danday* in both the active and middle voices.

While further work is necessary to determine the exact function of *-a* in Kharia, the following summary would seem to adequately describe its more general functions. It is important to stress that we are only dealing here with tendencies, not hard-and-fast rules:

- It is only found in Sadri loan words;
- It denotes that the contentive morpheme usually has, if it does not in fact require, an animate subject. With the possible exception of *bhunbhunay* ‘buzz (of bees)’, it also appears that a certain amount of reflective capacity, especially with respect to volitionality, is required by this subject, even if it is not human, although there remain a few exceptions, such as *gajgajay* ‘overflow’;
- On the other hand, the non *a*-marked form, i.e., a form in *-e*, usually does not have an animate subject;
- Where the *-e/-ay* contrast is found, the form in *-ay* generally occurs with active markers, the form in *-e* with those of the middle voice;
- While these traits are compatible with a causative (or perhaps transitivizing) interpretation of *-a* in many cases, it is not a causative or transitive marker *per se*, as is clearly shown by the fact that many forms ending in *-ay* can also be used intransitively.

Also, as we shall see below in Section 6.4.2.2, some speakers exploit the distinction between *-e* and *-ay* in morphemes which have both forms in a similar fashion to the use of the middle and active, with which *-e* and *-ay* are closely connected.

“Light verbs” or “conjunct verbs”

A number of predicates have semantic bases which consist of a contentive morpheme and a kind of “auxiliary verb”, generally referred to in linguistic studies as “light-verbs” or, in the Indological tradition, as “conjunct verbs”

The large majority of examples of this construction have been imported wholesale from Indo-Aryan (cf. Malhotra, 1982: 216), and the most common “light verbs” by far are *karay* ‘make, do’ and *la?* ‘EMOT’, both from Indo-Aryan.⁹

⁹ Or, depending on one’s definition of “light verb”, also *hoy* ‘be / become’, also of Indo-Aryan origin.

31. *kornis karay* 'attempt'
attempt do
32. *khatam karay* 'finish'
end do
33. *kheti karay* 'practice agriculture'
agriculture do

For examples of *la?*, see Section 6.3.4.2, where a number of experiential predicates of this type are discussed.

Not coincidentally, almost all of the contentive morphemes appearing with the "light verbs" from Indo-Aryan themselves derive from Indo-Aryan, where this construction is an extremely productive means of forming predicates from adjectives and nouns and also for accomodating foreign elements as predicates. There can be no doubt that Kharia has borrowed this use from Indo-Aryan, especially Sadri, as there is in fact no "need" for it in Kharia, since as we saw in Chapter 4, any contentive morpheme in Kharia can be used predicatively.

This also holds for the three examples given above. Cf.: *kornis* '(ACTIVE) try', *khatam* '(MIDDLE) come to an end' / '(ACTIVE) bring to an end, finish (TR)', *kheti* '(ACTIVE/MIDDLE) practice agriculture' (no apparent semantic difference).

(More or less) Fixed expressions?

Finally, there are a number of examples in this author's corpus and in Pinnow's texts (1965a, b) in which the complex semantic base would seem to be a more-or-less fixed expression. Consider the following example from a conversation:

34. *am=bar pe? beso=ki=bar?*
2=2HON rice good=MID.PST=2HON
'Have you eaten enough?'

However, the following two examples, from this author's own corpus, suggest that speakers do have a certain amount of freedom in constructing such predicates, where the semantic relations between the various contentive morphemes are left implicit:

35. *yo=na yona ethe²d god=ki bhere modi=ya?*
 see=INF REP annoy C:TEL=MID.PST time Modi=GEN
ma=dom timson sonol gon²ti²n=na col=ki.
 mother=3POSS fire firewood cook.rice=INF go=MID.PST
 ‘When she got annoyed at watching [them fight], Modi’s mother
 went to gather firewood for cooking (= she went to *fire-firewood-
 cook rice*).’ [RD, 2:55]
36. *boto²n=ta=pe ho=ki lutui su=kon pe? cakhna?*
 fear=MID.PRS=2PL that=PL clothes put.ON=SEQ rice curry
i²j²thaj kinbhar=na=pe, in=ko la²e=na=in.
 cow.dung courtyard=MID.IRR=2PL 1SG=CNTR fight=MID.IRR=1SG
 ‘Those of you who are afraid, you put on your [house] clothes and
 see to your house work like cooking and cleaning the courtyard with
 cowdung, but I will fight.’ [Kerke²t²ā, 1990: 7]

6.3.4 Experiential predicates / “psyche verbs”

Experiential predicates in Kharia follow one of two basic patterns: Either the experiencer appears in the direct case and is the subject of the construction, or this element appears in the oblique case, marked by =*te*, and the stimulus is the subject. This second pattern is the Kharia equivalent of the so-called “dative-subject” construction found in most South Asian languages.

6.3.4.1 Experiencer as subject

The predicate often consists of a simple lexical base, such as e.g. *beto²d* ‘hunger; hungry; become hungry’, or *ba²j* ‘like’, followed by TAM and PERS/NUM/HON markers:

37. *beto²d=si²d=in* ‘I am (= have become) hungry.’
 hunger=PERF=1SG
38. *in ho=ka²t=te ba²j=ta=in.* ‘I like him/her.’
 1SG that=SG.HUM=OBL like=MID.PRS=1SG

The predicate may also be of the “light verb” type (a so-called “conjunct verb”), discussed in Section 6.3.3. In this case, depending on the choice of the “light verb”, either the experiencer or the stimulus can be the sub-

ject. Light verbs typical of (possibly) agentive predicates, such as *karay* 'do' or *un* 'put, place', have the experiencer as the subject and the contentive morpheme referring to the psychological state as the object. In such cases, the stimulus is usually marked for the oblique case.

39. *i ponomosor ijn=te janmo alar dular karay=te?*
 q God 1sg=OBL always love love do=ACT.PRS
 'Does God always love me?' [HJPa:104]

Alternatively, the stimulus appears in the genitive:

40. *la? purkha=ya? yad karay[=e]=niŋ*
 then ancestor=GEN memory do=MID.IRR=1PL.INCL
 'Then we should remember the the ancestors.' [MS, 1:161]

6.3.4.2 Oblique-case experiencers

If the "light verb" is one of the following, the stimulus is the "subject" and the experiencer appears in the oblique case (given here with their respective meanings in predicative function): *la?* 'seem; attach; EMOT', *dho?* 'grab', *jaʔb* 'catch' (with sickness or insanity, where the ailment "grabs" or "catches" the experiencer), *qel* 'come', *aw* 'remain', *hoy* 'become' and *ayiʔj* 'QUAL.PRS'. The same case-marking pattern is found with *kosu* 'hurt', the causative of *kosu* 'be hurt, injured'

41. *ijn=te musa jeri kosu=tej ro kulqaʔ=jo*
 1sg=OBL today body hurt-<CAUS>=ACT.PROG and fever=ADD
jab=si?
 catch=PERF
 'Today my body is hurting me and fever has also taken hold of (= caught) me.' [BB, 2:23]

42. *aʔb=ko kulu=te koroʔb um aw-mon laʔ=ki.*¹⁰
 now=CNTR turtle=OBL quiet NEG remain-mind EMOT=MID.PST
 'But now the turtle did not wish to remain quiet.' [TK, 1:60]

¹⁰ Cf. Section 4.6.2.1 on *-mon* 'mind'

A number of experiential predicates can be of both types above:

43. Direct-case experiencer	Oblique-case experiencer	Meaning
<i>beto²d</i>	<i>beto²d la?</i>	‘become hungry’
<i>kulqa?</i>	<i>kulqa? ja²b</i>	‘have / get a fever’
<i>raŋga</i>	<i>raŋga la?</i>	‘feel cold’
<i>phaham / phaham karay</i>	<i>phaham aw / dɛl</i>	‘remember’

6.3.4.3 Overview of experiential predicates

Table 6.10 provides an overview of experiential predicates in Kharia. Note that the most common “light verb” is *la?* ‘seem; EMOT’. All others are far less common, such as *un* ‘place’, of Kharia origin.

The table is by no means exhaustive, and an empty slot does not necessarily imply that the corresponding form is not found. Rather, it only means that it was not found in our data.

Table 6.10: Experiential Predicates in Kharia

<i>Semantic Area (roughly)</i>	<i>Experiencer = Subject</i>	<i>Stimulus = Subject</i>
Hunger and Thirst	<i>beto²d</i> ‘become hungry’	<i>beto²d la?</i> / <i>hoy</i> ‘(same)’
	<i>beto²d qa?</i> ‘become thirsty’	<i>beto²d qa? la?</i> ‘(same)’
Sickness, Insanity / Bodily Sensations		<i>qa? piyas la?</i> ‘become thirsty’
	<i>bimar</i> ‘become sick’	<i>bay ja²b</i> ‘go crazy’
		<i>du?kho la?</i> ‘feel pain, suffering, sorrow; be unhappy’
		<i>giniŋ la?</i> ‘feel hot’ (of people)’
	<i>kosu</i> ‘hurt’ (TR)	<i>kosu la?</i> ‘be sad; hurt (IRR)’
	<i>kulqa?</i> ‘have / get a fever’	<i>kosu dho?</i> ‘become sick’
	<i>mo²dqa?</i> ‘have / get a cold’	<i>kulqa? ja²b</i> ‘(same)’
	<i>raŋga</i> ‘feel cold’	<i>mo²dqa? dho?</i> ‘of the eyes to water’
		<i>raŋga la?</i> ‘(same)’

Table 6.10 (cont.)

<i>Semantic Area</i> (roughly)	<i>Experiencer = Subject</i>	<i>Stimulus = Subject</i>
Joy, Sorrow, Anger	<i>du?kxo hoy</i> 'become unhappy'	<i>baru la?</i> 'be happy, like' <i>du?kxo la?</i> 'feel pain, suffering, sorrow; become unhappy'
	<i>horo?j</i> 'become angry (at s.o.)'	<i>kosu la?</i> 'be sad; hurt'
	<i>kharab</i> 'become unhappy'	
	<i>khus</i> 'become happy'	
	<i>lere?</i> 'become happy, rejoice'	<i>lere? ayi?j / hoy</i> 'be(come) happy'
	<i>rase</i> 'become happy' <i>su?kxo</i> 'become happy'	
Liking / Loving and Disliking / Hating		<i>accha la?</i> 'like'
	<i>alar</i> 'love'	<i>alar dular un</i> 'like' (experiencer marked by <i>buŋ</i> 'INST')
	<i>ba?j</i> 'like'	<i>ba?j-ba?j hoy</i> '(same)' <i>baru la?</i> 'like' <i>cahi</i> 'love'
	<i>dular</i> 'like, love'	
	<i>dular karay / un</i> 'love'	
	<i>dular alar karay</i> 'love'	
	<i>dularay</i> 'love'	
	<i>gana?j karay</i> 'hate'	
	<i>ghinay</i> 'hate'	
	<i>isar</i> 'hate'	
	<i>khus un</i> 'like' (cf. <i>khus</i> 'become happy')	
	<i>lebui</i> 'love'	
<i>lebui un</i> 'love'	<i>lebui la?</i> 'feel pity' <i>majur ayi?j / la?</i> 'like'	
<i>mala?</i> 'like'		
<i>lam</i> 'want'	<i>X-mon la?</i> 'want to X, feel like X-ing'	

Table 6.10 (*cont.*)

<i>Semantic Area</i> (<i>roughly</i>)	<i>Experiencer = Subject</i>	<i>Stimulus = Subject</i>
Memory, Thought, Interest	<i>konoʔd</i> ‘think (up); remember’	
	<i>khunbher</i> ‘remember’	
	<i>p(h)aham, p(h)aham karay</i> ‘remember’	<i>p(h)aham aw / dɛl</i> ‘(same)’
	<i>yad(gari) karay / un</i> ‘remember’	<i>yad ayiʔj / dɛl</i> ‘(same)’
	<i>khyal ayiʔj / un</i> ‘be of the opinion’	
		<i>ruci dhoʔ</i> ‘become interested’
	<i>socay (un)</i> ‘think’	
Hope, Fear, Trust, etc.	<i>asra, asra un</i> ‘hope’	<i>asra ayiʔj</i> ‘(same)’
	<i>botoŋ</i> ‘fear’, <i>botoŋ</i>	<i>botoŋ laʔ</i> ‘(same)’
	<i>saphay</i> (same)	
	<i>bhorsa un</i> ‘trust’	
	<i>cette</i> ‘worry’	<i>cinta hoy</i> ‘worry’ <i>sog laʔ</i> ‘pity’

6.3.5 *Causative and reciprocal marking*

The causative marker is a derivational affix which combines with an underived contentive morpheme—as a prefix, an infix, or both in double causatives—in an often rather idiosyncratic way to form a new (derived) morpheme. The reciprocal on the other hand is a phonological word which appears before the contentive morpheme which it modifies. These two markers will be dealt with in this section as they form part of the semantic base.

Causative: //OʔB//

The causative increases the valency of the contentive morpheme by one by introducing a higher or superordinate agent, i.e., causing someone to carry out an action or causing an otherwise avolitional spontaneous event to happen. Except for the generic function of the middle voice (6.4.2.2) and unless the semantic base is simultaneously marked for the reciprocal or the passive/reflexive (6.5.1), a TAM/PERSON-syntagma marked for the causative always appears in the active voice.

With monosyllabic contentive morphemes, the causative marker appears as a prefix, with bisyllabic morphemes, it is generally an infix, although most speakers also accepted the use of the causative marker as a prefix with bisyllabic morphemes as well. The exact distribution of the various allomorphs is partially idiosyncratic and there is a considerable amount of speaker-specific variation. There are nevertheless clear tendencies:

- As an infix, the causative marker is generally realized as either *-b-* or *-ʔb-*:

44. <i>halo</i> 'get wet'	CAUSATIVE: <i>ha<ʔb>lo</i> 'make wet'
<i>jereʔd</i> 'burn (ITR) (of rice in pot)'	<i>je<ʔb>reʔd</i> 'burn (TR) (rice in a pot)'
<i>juda</i> 'become separate'	<i>juda</i> 'separate' (TR)

Before labials, and for many speakers elsewhere as well, the infix is realized as *-ʔ-*:

45. <i>botoŋ</i> 'become frightened'	CAUSATIVE: <i>bo<ʔ>toŋ</i> , <i>bo<ʔb>toŋ</i> 'scare'
<i>bhore</i> 'become full'	<i>bho<ʔ>re</i> 'fill up' (TR)
<i>jhaloŋ</i> 'become tall, high'	<i>jha<ʔ>loŋ</i> 'lengthen'
<i>kabuŋ</i> 'crawl'	<i>ka<ʔ>buŋ</i> 'help crawl'
<i>lemeʔd</i> 'go to bed; go to sleep'	<i>le<ʔ>meʔd</i> 'put to bed'
<i>jimi</i> 'become named'	<i>ji<ʔ>mi</i> 'name'
<i>sebol</i> 'become sweet'	<i>se<ʔ>bol</i> 'sweeten (TR), make sweet'

- The form of the prefix is generally *oʔb-* before most consonants and *ob-* before vowels. Some speakers prefer the form *oʔ-* over *oʔb-*. This is the usual form of the causative with monosyllabic contentive morphemes and, at least in interviews, was also often cited for bisyllabic contentive morphemes as well (in addition to the infix form, which was always accepted with bisyllabic morphemes). In interviews, the form *oʔb-* was generally given even if the contentive morpheme began with a vowel, whereas in actual speech it appears to generally be realized as *ob-* [ɔb] before vowels. As many of the examples of causatives with this prefix are from texts, where pre-glottalized consonants were not always indicated, the forms given here do not always give the actual

pronunciation. As pre-glottalization is not phonemically distinctive, however, this causes no ambiguities.

46. *gaʔj* ‘fry (food)’ CAUSATIVE: *oʔ-gaʔj* ‘cause to fry,
have s.o. fry’
noʔ ‘eat’ *oʔb-noʔ* ‘feed, make
s.o. eat’
puʔd ‘jump’ *ob-puʔd* ‘cause to jump’
raʔ ‘blossom’ *ob-raʔ* ‘cause to
blossom’
yo ‘see’ *ob-yo* ‘show’

– A small number of contentive morphemes take the causative with the form *o-*. Preceding a vowel, *-y-* is inserted to avoid a hiatus (2.4).

47. *dam* ‘arrive’ CAUSATIVE: *o-dam* ‘cause to arrive;
bring’
deʔb ‘ascend’ *o-deʔb* ‘raise; offer up
in sacrifice’
duʔj ‘bend’ (TR (ACT) / ITR (MID)) *o-duʔj* ‘bend’ (TR,
same as active)
ej ‘return’ (ITR) *o-y-ej*, *oʔb-ej* ‘bring
back’
geʔb ‘burn’ (ITR) *o-geʔb*, *ob-geʔb*
‘burn’ (TR)

– There is also a relatively small number of slightly irregular causative forms, occasionally with a somewhat unpredictable meaning. For example, the causative form of *hoy* ‘become’, which is morphologically either regular, *ob-hoy*, or slightly irregular, *ho<bh>oy*, does not mean ‘cause to become’ or ‘create’ but rather ‘distribute’

The following list presents five additional slightly irregular forms, but all with predictable meaning.

48. *aw* ‘live, stay, remain’ CAUSATIVE: *obh-aw* ‘cause to live,
stay, remain’
buŋ ‘finish (ITR), come to an end’ *om-boŋ* ‘finish’ (TR)
diyā ‘enter’ *di<ʔbh>ar* ‘bring in,
cause to enter’

<i>goŋ</i> 'cook (rice) (TR)'	<i>oʔ-guŋ</i> (but double causative: <i>oʔb-go<ʔ>ŋ</i>) 'cause to / have s.o. cook (rice)'
<i>gur</i> 'fall'	<i>u-gur, ub-gur, o-gur, oʔ-gur, ob-gur</i> 'drop'

Double causatives can be formed from most bisyllabic contentive morphemes. Whether or not a double causative occurs (or was accepted in interviews) seems to depend on whether or not the situation it denotes occurs with a certain regularity, and opinions occasionally differed to some extent during interviews on whether or not a particular double causative was acceptable. As there is only one example of a double causative in our entire corpus (*oʔb-kho<ʔb>da* [CAUS-paint.oneself-<CAUS>] 'have s.o. paint s.o. else'), all of the information in the following is from interviews with native speakers.

Double causatives of bisyllabic contentive morphemes are formed by the presence of both the infixal and prefixal form of the general causativizer. Table 6.11 provides a few examples.

Table 6.11: Causatives and double causatives

<i>Underlying contentive morpheme</i>	<i>Simple causative</i>	<i>Double causative</i>
<i>ajoʔd</i> 'dry up (TR)'	<i>a<ʔb>joʔd</i> 'dry up (TR)'	<i>oʔb-a<ʔb>joʔd</i> 'have s.o. dry up'
<i>aloŋ</i> 'sing'	<i>a<ʔb>loŋ</i> 'have s.o. sing'	<i>ob-a<ʔb>loŋ</i> 'have s.o. make s.o. sing'
<i>are</i> 'descend'	<i>a<ʔ>re, a<ʔb>re</i> 'lower' (TR)	<i>oʔb-a<ʔ>re</i> 'have s.o. lower'
<i>lemeʔd</i> 'go to bed'	<i>le<ʔ>meʔd</i> 'put s.o. to bed'	<i>oʔb-le<ʔ>meʔd</i> 'have s.o. put s.o. to bed'
<i>sore</i> 'become ready'	<i>so<ʔb>re</i> 'prepare'	<i>ob-so<ʔb>re</i> 'have s.o. prepare'

With monosyllabic contentive morphemes whose causative form takes the prefix *o-*, double causatives can be formed with the prefix *oʔb-*, which functions here as a double causative marker. This does not apply to those contentive morphemes which have alternate causative forms in *o-* / *oʔb-*: *eŋ* 'return' (ITR), CAUSATIVE: *o-y-eŋ* / *oʔb-eŋ* 'return' (TR), NO DOUBLE CAUSATIVE or *geʔb* 'burn' (ITR), CAUSATIVE: *o-geʔb* / *oʔb-geʔb* 'burn' (TR), NO

DOUBLE CAUSATIVE. Table 6.12 provides two examples of this double causative formation.

Table 6.12: Causatives and double causatives in *o-* / *oʔb-*

<i>Underlying contentive morpheme</i>	<i>Simple causative</i>	<i>Double causative</i>
<i>qam</i> 'arrive'	<i>o-qam</i> 'bring' (= cause to arrive)	<i>oʔb-qam</i> 'have s.o. bring'
<i>qeʔb</i> 'rise; climb'	<i>o-qeʔb</i> 'raise; offer up; sacrifice'	<i>oʔb-deʔb</i> 'have s.o. sacrifice'

The rules stated above would seem to suggest that, with the exception of the extremely small group of contentive morphemes which have the causative / double causative alternation in *o-/oʔb-*, only bisyllabic contentive morphemes should be expected to have double causative forms, as only these morphemes can take the infix <*b*>. However, speakers in interviews did accept double causatives of monosyllabic morphemes. Table 6.13 presents a few examples.

Table 6.13: Causatives and double causatives of monosyllabic contentive morphemes

<i>Underlying contentive morpheme</i>	<i>Simple causative</i>	<i>Double causative</i>
<i>qal</i> 'cover (a pot)'	<i>ob-qal</i> 'have s.o. cover'	<i>oʔb-qa<ʔb>l</i> 'have s.o. make s.o. cover'
<i>qoʔq</i> 'take'	<i>oʔb-qoʔq</i> 'have s.o. take'	<i>oʔb-qo<ʔb>q</i> 'have s.o. make s.o. take'
<i>gil</i> 'hit'	<i>oʔb-gil</i> 'have s.o. hit'	<i>oʔb-gi<ʔ>l</i> 'have s.o. make s.o. hit'

This may be due to the interview situation in which this information was elicited, as the simple causative of monosyllabic contentive morphemes was also occasionally cited with the infix, e.g., *qal* 'cover (a pot)', causative *ob-qal* or *qa<ʔb>l*, although there are no naturally occurring examples of this in the corpus. To what extent this is a naturally occurring construction or due to the interview situation awaits further study.¹¹

¹¹ It seems reasonable to assume that this is only possible before enclitics beginning with a vowel such as =*oʔ* 'ACT.PST', so that the strict (C)V(C) structure of the syllable is not violated.

Occasionally, we also come across forms marked as double causatives but with the same meaning as the simple causative. This is quite rare and the examples given in Table 6.14 are the only ones found in our data.

Table 6.14: Morphological double causatives with simple causative semantics

<i>Underlying contentive morpheme</i>	<i>Simple causative</i>	<i>Morphological double causative, semantic simple causative</i>
<i>katib</i> 'collect, assemble' (TR)	<i>ka<ʔb>tib</i> , <i>oʔb-katib</i> 'gather' (TR)	<i>oʔb-ka<ʔb>tib</i> 'gather (TR)'
<i>socay</i> 'think'	<i>oʔb-socay</i> , <i>so<ʔb>cay</i> 'cause to think'	<i>oʔb-so<ʔb>cay</i> 'have s.o. think'
<i>yar</i> 'flee'	<i>oʔb-yar</i> 'chase away, make flee'	<i>oʔb-ya<ʔ>r</i> 'chase away, make flee'

Finally, contentive morphemes which have been borrowed from Indo-Aryan and which end in *-ay* often form their causatives by inserting *-(u)w-* before *-ay*. This may occur either with or without the usual causative prefix *oʔb-*.

49. *akbakay* 'make nervous' CAUSATIVE: *akbak<uw>ay*
bhosṭay 'knock over' *oʔb-bhosṭ<uw>ay*
gohray 'ask for forgiveness' *oʔ-gohar<w>ay*

The status of this construction is somewhat enigmatic. For some speakers (in interviews), the inclusion of the causative prefix *oʔb-* is facultative and had no semantic effect, and the infix *-uw-* was the causative marker. For other speakers, both the prefix *oʔb-* and the infix *-uw-* together expressed the causative and the prefix was not facultative (cf. also Malhotra, 1982: 170–171). For others still, the infix *-uw-* expressed the simple causative and the inclusion of the prefix *oʔb-* denoted a double causative. Whether these differences are due to the interview situation, personal preferences or regional dialects is unclear.

With morphologically (and semantically) simple causatives, the highest agent is the subject in the active and the intermediate agent, the one who is to perform the action, appears either in the oblique case or may be followed by either of the postpositions *buṅ* 'INST' or *tay* 'ABL' (cf. also Malhotra, 1982: 164).

50. *ij̄ ho=kaṛ=te / ho=kaṛ buṅ dura*
 1sg that=SG.HUM=OBL that=SG.HUM INST door

ob-rukḥ=oʔj̄.

CAUS-open=ACT.PST.1SG

‘I had him / her open the door.’

51. *ij̄ dhangar beṭa tay kamu ob-karay=oʔ[j̄].*
 1sg servant boy ABL work CAUS-do=ACT.PST.1SG

‘I made the servant boy do the work.’ [Malhotra, 1982: 170]

In the (semantic and morphological) double causative, the higher intermediate agent is always marked in our data by the postposition *buṅ* ‘INST’ and the lower intermediate agent is marked by the oblique case.

52. *ij̄ am=aʔ buṅ ho=kaṛ=te biru tay*
 1sg 2SG=GEN INST that=SG.HUM=OBL mountain ABL

oʔb-a<ʔ>re=yoʔj̄.

CAUS-descend-<CAUS>=ACT.PST.1SG

‘I had you make him bring [something] down from the mountain.’

As the following examples show, causative marking in TAM/PERSON-syn-
tagmas with complex semantic bases attaches only to the semantic sub-
bases it has scope over:

53. *musniṅ jhaṛi kopuṛuʔ=ki jume katiʔb=kon socay=oʔ=ki*
 one.day all man=PL assemble gather=SEQ think=ACT.PST=PL

no “ho dano=te i=gud=ga tar oʔ-gur=e=niṅ?”

CMPL that demon=OBL what=like=FOC kill CAUS-fall=ACT.IRR=1PL.EXCL

‘One day, all the men came together (= ‘having assembled [and] gathered’) and thought “How will we kill (= ‘kill [and] cause to fall’) that demon?”’ [MT, 1:66]

54. *yohan ho=ki tiʔj̄ gam o-y-eṅ=oʔ no:*

John that=PL side say CAUS-y-return=ACT.PST CMPL

‘John answered them, saying (= *say-returned*)’ [HJPa:88, ln. 26]

“*Reciprocal*”: *kol*

Of the four major situation types discussed in Lichtenberk (1985) for reciprocal constructions, namely reciprocal, reflexive, collective and chaining situations, *kol* in Kharia can express three of these: reciprocals, collectives and, to a much lesser extent, chaining situations. However, its use is not restricted to these three types.

The marker *kol* generally has a reciprocal interpretation, i.e., two entities mutually affect one another in some manner, or to quote Lichtenberk (1985: 21) “The reciprocal situation can then be defined as one in which there are two participants, A and B, and the relation in which A stands to B is the same as that in which B stands to A . ” The following presents a few examples.

55. *kol baʔj=ki=kiyar*
 REC like=MID.PST=DU
 ‘they (DU) liked each other’

56. *kol khed=ki=kiyar*
 REC bite=MID.PST=DU
 ‘they (DU) bit each other’

57. *kol enduʔ=ki=kiyar*
 REC stare=MID.PST=DU
 ‘they (DU) stared at each other’

With many predicates *kol* can also denote that the two entities did something together, as the following forms show. This is referred to by Lichtenberk (and others) as “collective”

58. *kol along=ki=kiyar*
 REC sing=MID.PST=DU
 ‘they (DU) sang together’

59. *kol lereʔ=ki=kiyar*
 REC rejoice=MID.PST=DU
 ‘they (DU) rejoiced together’

60. *kol berod=ki=kiyar*
 REC stand.up=MID.PST=DU
 ‘they (DU) stood up together’

61. *kol buli=ta²j=kiyar*
 REC wander=MID.PROG=DU
 ‘they’re (DU) walking together’

The third situation type discussed in Lichtenberk (1985) which is also covered by *kol* in Kharia is the “chaining” situation, exemplified by the following types:

62. *kol buli²j=ta²j=kiyar* ‘they’re (DU) following each other’
 REC follow=MID.PROG=DU
63. *kol kuday=ki=kiyar* ‘they (DU) chased each other’
 REC chase=MID.PST=DU
 (“closed chain”, Lichtenberk, 1985: 25)

The basic meaning of this category appears to be that of mutual affect-
 edness in general, and the use of *kol* is not restricted to the three types
 mentioned above. That is, both (or more) entities equally carry out and
 are equally affected by an action. This would then explain the usage in the
 following, where *kol* refers to an object owned by the subject:

64. *kol bagray=ki=kiyar* ‘they (DU) destroyed something they own.’
 REC destroy=MID.PST=DU
65. *ho=kiyar pothi=te kol lusu=ki=kiyar* ‘They (DU) grabbed
 that=DU book=OBL REC steal=MID.PST=DU each other’s book.’

Similarly, *kol* can be used to signal that the event was carried out to the
 mutual benefit of those involved.

66. *kol bay=ki=kiyar* ‘they (DU) built for each other’
 REC make=MID.PST=DU
67. *kol lam=ki=kiyar* ‘they (DU) searched for each other’ (with a
 REC search=MID.PST=DU beneficial reading)

Closely connected to the idea of mutual benefit is that of helping one another, which *kol* can also denote:

68. *kol absi?b=ki=kiyar* 'they (DU) helped each other start'
REC start=MID.PST=DU

69. *kol mu?=ki=kiyar* 'they (DU) helped each other come out'
REC emerge=MID.PST=DU

A number of forms allow for more than one interpretation, as in the following example, where both a reciprocal and a collective interpretation are possible.

70. *kol baŋce=ki=kiyar* 'they (DU) were saved together / they (DU)
REC save=MID.PST=DU saved each other'

71. *kol araŋ=ki=kiyar*
REC pour.out.rice=MID.PST=DU
'they (DU) poured out each other's rice' / 'they (DU) poured out the rice for each other'

kol has no independent lexical meaning. A TAM/PERSON-syntagma marked for the reciprocal always appears in the middle voice. With respect to the mobility of *kol*, see Section 4.6.4.

6.4 TAM / BASIC VOICE MARKING

6.4.1 *The basic TAM categories*

Almost all morphologically fully finite TAM/PERSON-syntagmas in Kharia are marked as either active or middle, while non-finite forms are neutral with respect to these two categories. These two categories are marked by portmanteau morphemes which simultaneously express four of the seven basic TAM categories, as well as basic voice (active, middle).

The remaining three categories, "Past II", the perfect and the optative, are neutral with respect to basic voice.¹² These forms are presented in

¹² With respect to the perfect, it is actually only the present perfect which is morphologically unmarked for tense and basic voice, except in habitual use (6.4.1.5). The past perfect

Table 6.15. In this section we will discuss these TAM categories in detail. The active and middle voices will be dealt with in 6.4.2.

Table 6.15: The basic TAM/voice categories

	<i>Middle</i>	<i>Active</i>
Past (PST) (aspectually neutral)	=ki	=(y)oʔ
Present General Imperfective (PRS)	=ta	=te
Present Progressive (PROG)	=taʔjɖ	=teʔjɖ
Irrealis (IRR)	=na	=e
“Past II” (PST.II)		=khoʔ
Perfect (PERF)		=siʔ(ɖ)
Optative (OPT)		guɖuʔ / guɖuʔ

6.4.1.1 *Past and “Past II”*

The past and Past II are used for events and states which have already occurred at the time of the speech-act. Before =oʔ, a morpheme-final consonant is devoiced and aspirated: Final /b/ becomes /pʰ/, /ɖ/ becomes /tʰ/, and /ŋ/ (< */g/) is realized as /kh/:

72. *jaʔb* ‘catch’ *japh=oʔ* ‘s/he caught’
konoʔɖ ‘think up’ *konoʔh=oʔ* ‘s/he thought up’
dhoʔ ‘grab’ *dhokh=oʔ* ‘s/he grabbed’

After a morpheme-final vowel, the active simple past form is realized as =yoʔ:

73. *yo=yoʔ* ‘s/he saw’ *thisa=yoʔ* ‘s/he called out’
 see=ACT.PST call.out=ACT.PST

Younger speakers, however, show a tendency to insert =yoʔ in all environments, also after morpheme-final consonants:

and irrealis perfect both mark for basic voice. The optative and Past II, on the other hand, never mark for basic voice.

74. *u gam=kon etwa beṭ=ḍom=te juda=na=? thoṅ*
 this say=SEQ Etwa son=3POSS=OBL separate-<CAUS>=INF=GEN PURP
khori=ya? pahan paṅc=ki=te ro mudh mudh
 village.section=GEN priest Panch=PL=OBL and special REP
lebu=ki=te paṅceiṭ=te rema?yo?
 man=PL=OBL panchayat=OBL call=ACT.PST
 'Having said this, Etwa *called* the village's priest, the Panch and
 [other] special people to the *panchayat* in order to get rid of (= "sepa-
 rate") his son.' [RD, 2:91]

Both the simple past and Past II are common in narrating chains of events in the past, such as the simple past in the following example:

75. *rel chu?te=na thoṅ pōga baje=ki la? bheṛi*
 train leave=INF PURP horn be.sounded=MID.PST then sheep
merom=ki ho=ti?j=ga col=ki=may.
 goat=PL that=side=FOC go=MID.PST=3PL
 'The horn *sounded* for the train to leave, so the sheep and goats *went*
 in that direction.' [RD, 1:11]

Although the two past tenses are compatible with terminated past events and are perhaps most often found in this environment, this is not necessarily so:

76. *ḍa? gim=o? (ro tam jou gim=te?j).*
 water rain=ACT.PST and now up.to rain=ACT.PROG
 'It *rained* (and is still raining).'

Thus, in contrast to the periphrastic imperfective (6.8.1.1), which is positively marked for imperfectivity, the simple past and Past II are both aspectually neutral.

The use of the two past tenses is also very common in narrating past states when "durativity" is not stressed:

77. *moṅ raja rani aw=ki=kiyar.*
 one king queen QUAL=MID.PST=DU
 'There *were* a king and queen.' [AK, 1:3, 4]

raja=ya? *jimi* *aw=ki* *sembho* *oɔo[ʔ]* *rani=ya?* *jimi*
 king=GEN name QUAL=MID.PST Sembho and queen=GEN name

aw=ki *dakay.*

QUAL=MID.PST Dakay

‘The king’s name *was* Sembho and the queen’s name *was* Dakay.’

The Past II marker =*kho?* derives from the marker of the past perfect, =*sikh=o?*, which always appears in the active. The Past II does not seem to differ at all semantically from the simple past, although the active/middle opposition is neutralized in this category, as the form is always =*kho?*.¹³ Its use is considered incorrect, although it is very common in the spoken language, especially among younger speakers.

78. *ho=te* *kheti uslo?* *bes bes* *aw=kho?*
 that=OBL(=‘there’) field soil good REP QUAL=PST.II

‘The fields and the soil there *were* good.’ [MT, 1:76]

The Past II would appear to be a rather recent form, although it is not certain how recent it is. The fact that it is quite commonly used by younger speakers and that elder speakers, at least those we questioned, tended to frown upon its use, could be taken as evidence that its appearance was very recent, but the evidence is not entirely conclusive.

To begin with, elder speakers we questioned on this issue were all highly educated, some of whom taught Kharia at the university. Their intuitions may thus reflect a prescriptive attitude towards this category.

Furthermore, although a check of the first half of the texts in Pinnow (1965a) turned up no occurrences of this morpheme, a closer look reveals that this is at least partly due to the editorial work which Pinnow undertook. For example, in Pinnow (1965a: 80, second line of Kharia text from the top and also fn. 6) we find the form *remakh=o?* [call=ACT. PST] ‘he called’ In the footnote, however, we find the comment “Text: *rema?kho?*” In other words, Pinnow “corrected” this form to the standard, prescriptive *remakho?* whereas the form in the text itself was the Past II form *rema?=kho?* ‘call=PST.II’ Thus, only a careful search of all of the footnotes in Pinnow’s (1965a;b) texts will reveal the degree to which

¹³ Alternatively, one could analyze this as the Past II marker =*kh* and the simple past active marker =*o?*. I will retain the analysis of this form as =*kho?*, unmarked for basic voice, as there is no corresponding middle form.

the Past II was already in use at that time and by which speakers but, as Pinnow's texts are meticulously edited, such an undertaking is certainly possible.

6.4.1.2 *Present general imperfective*

The present general imperfective is used for actions which occur on a regular basis, hold at the moment of utterance, as a "historical present" or for imminent, intentional future actions (at least in the first persons, including adhortatives).

The present active marker =*te* is regularly realized as =*t* before the first and second-person, singular markers =*ɨn* and =*em*, respectively. To signal that this is the active and not the middle marker in such environments, the marker will be given as =*t[e]* in these cases:

79. *oʔ-gur=t[e]=ɨn* 'I drop [regularly]'
CAUS-fall=ACT.PRS=1SG

Similarly, after the present middle marker =*ta*, the /i/ in =*ɨn* and the /e/ in =*em* may be deleted, although this is not obligatory:

80. *col=ta=n* 'I will go' but *qhalanje=ta=ɨn* 'I will take a nap'
go=MID.PRS=1SG take.a.nap=MID.PRS=1SG

The following two examples show the use of the present tense for events which occur on a regular basis. In (81), the speaker is telling us about the cooking habits of the Kharia:

81. *khori poʔda boʔ=ki=te tam jou kharɨya=ki*
village.section village place=PL=OBL now up.to Kharia=PL
kaqom cakhna hinte sanjan umay may=te. ho=ki
fish curry LOC turmeric NEG.3PL mix=ACT.PRS that=PL
tenton=ga may=te=ki.
tamarind=FOC mix=ACT.PRS=PL

'Up to the present day, in the villages and village sections, the Kharia do not mix turmeric into fish curries. They mix in tamarind.'

In the following example, the speaker is explaining the derivation of the name of the month of March (*qokdaʔ*) as deriving from the name of Queen Dakay (*dakay*):

82. *u ḍakay rani=ya? jimi buḥ=ga, ikon, marc mahina=te*
 this Dakay queen=GEN name INST=FOC umh March month=OBL
ḍokḍa? gam=te=ki.
 Dokda? say=ACT.PRS=PL
 ‘Because of this queen’s name, umh, *they call* the month of March
 “Dokda?”’ [AK, 1:5]

There is also at least one attested example in which a TAM/PERSON-syntagma in the present tense follows the same TAM/PERSON-syntagma marked for the irrealis, with an iterative reading:

83. *u=ki=te gam[=e]=iḥ gam=t[e]=iḥ, muda iḥ=ga*
 3=PL=OBL say=ACT.IRR=1SG say=ACT.PRS=1SG but 1SG=FOC
aṭke go²ḍ=ta=iḥ.
 be.stopped C:TEL=MID.PRS=1SG
 ‘I *tell* them *over and over*, but I am always stopped.’
 [Kerkeṭṭā, 1990: 3]

In its use as a “historical present”, the present often alternates freely with the past tense, even within a single sentence:

84. *la? porha=jo col=ki ro yo=te, ho maṛa boṭte*
 then Porha=ADD go=MID.PST and see=ACT.PRS that cave LOC
ḍam=ki ḍa?=te yo=yo?
 arrive=MID.PST water=OBL see=ACT.PST
 ‘Then Porha also went and *looks*, [he] arrived at the cave [and] saw
 the water.’ [AK, 1:39]

85. *ro kiṛo?=te lam=na col=ki. lam=na lamna=te solo?*
 and tiger=OBL search=INF go=MID.PST search=INF REP=OBL dog
kiṛo[?]=te kuy=o? ro gam=te “i=ghay ayi?jḍ=em
 tiger=OBL find=ACT.PST and say=ACT.PRS what=way QUAL.PRS=2SG
bhai, baru=ga no?”
 brother good=FOC Q
 ‘And he went to look for the tiger. Searching and searching he found
 the tiger and *says* “How are you brother, OK?”’ [BB, 3:45]

The following examples illustrate the use of the present for future events:

86. *kongher rag badli=kon gam=o? "ga?j=chi?q=iŋ,*
 boy voice change=SEQ say=ACT.PST fry=PERF=1SG
yo, ol kay=t[e]=iŋ?"
 mother.VOC bring BEN=ACT.PRS=1SG
 'The boy, changing his voice, said "I have fried [the meat], mother,
shall I bring some for you?"' [BB, 1:80]

87. *col=ta=ŋ ho=ki=ya? tay ŋo?=na thoŋ bor*
 go=MID.PRS=1SG that=PL=GEN ABL eat=INF PURP ask.for
ol[=e]=iŋ.
 bring=ACT.IRR=1SG
 'I will go beg for something (= *will go*, will ask for [and] bring) to
 eat from them.' [TK, 2:32]

It is also found with an adhortative sense:

88. *ani jahā ti?j col=ta=niŋ!*
 come.on INDEF.NHUM side go=MID.PRS=1PL.INCL
 'Come on! *Let's go* somewhere!' [TK, 1:7]

It could thus be argued that what is termed here the present general imperfective is more appropriately termed "nonpast", as this has been referred to in previous studies on Kharia (e.g., Peterson, 2002a, 2007). However, as this function of the present seems to be restricted to first persons, whereas the use of the irrealis for future time is found in conjunction with all persons, this category is considered a present tense here and examples such as (86)–(88) can be explained by assuming that they are considered more immediate and—as the subject is a first person—more definite, similar to the progressive (see below), whereas the use of the irrealis in these environments is unmarked in these two respects.

6.4.1.3 *Present progressive*

The progressive is primarily used for events and situations which hold at the time of utterance (89). Its marking consists of the present marker =*te/=ta* plus what appears to be a marker deriving from the qualitative

predicative marker *ayiʔjd*: *ayiʔjd* is realized as *ayiʔj* in all persons other than the first and second persons, singular. Similarly, the progressive markers *=teʔjd* (ACTIVE) and *=taʔjd* (MIDDLE) are realized as *=teʔj* and *=taʔj* in these environments, respectively.¹⁴

89. *in=te* *musa jeri* *kosu=teʔj* *ro kuldaʔ=jo*
 1SG=OBL today body get.sick-<CAUS>=ACT.PROG and fever=ADD
jab=siʔ.
 catch=PERF
 ‘Today [my] body *is hurting* me and fever has also taken hold of
 (= caught) me.’ [BB, 2:23]

Similar to the present tense, the progressive is also found with future events which are considered imminent and definite, (90)–(91), at least in the first persons, and in adhortatives.

90. (The witch said:) *in gone cokhay=ke del=taʔjd=in=ga*.
 1SG tooth sharpen=SEQ come=MID.PROG=1SG=FOC
 ‘I’ll come right back after I sharpen my teeth.’
 [BB, 1:65]

91. *am khis loʔ=na=m* *ele* *col=taʔj=le*.
 2SG angry CONT=ACT.IRR=2SG 1PL.EXCL go=ACT.PROG=1PL.EXCL
 ‘If you keep on getting angry, we will leave.’

As with the present tense, the progressive may also be used as a kind of “historical present”:

92. *laʔ ho=te* *khali khariya bheir aw=ki=may*
 then that=OBL(= ‘there’) only Kharia much live=MID.PST=3PL
oqoʔ hoqom jait=ko *mane ho=ki kuday=ga*
 and other ethnic.group=CNTR umh that=PL chase=FOC

¹⁴ Unlike Biligiri (1965: 58), I do not consider the segment /j/, or possibly /jd/, in *=tejd* / *=tjd* to be a separate marker. Although this does have a certain appeal to it, and although it is clearly the presence or absence of this element which determines the interpretation as progressive or non-progressive (“present immediate” in Biligiri’s terminology), I do not analyze this form as a separate morpheme as it can only appear in the present tense. Instead, I prefer to analyze these two morphemes merely as *=tej(d)* and *=taj(d)*, respectively, although this is admittedly a somewhat arbitrary decision.

ol=teʔj=ki.

v2:bring=ACT.PROG=PL

'Then only Kharia lived there and *they drive away* the other ethnic groups.'
[MS, 1:153]

The progressive is restricted to these "present" contexts, i.e., present event, historical present, definite/immediate future actions (first persons) and adhortatives. For past, future and hypothetical progressive meanings, the periphrastic imperfective is used (6.8.1.1).

6.4.1.4 *Irrealis*

The irrealis is used for future actions and states (93)–(94), in the imperative (95)–(96), and in conditional and counterfactual clauses (7.5.1.4). Note that when an imperative sense is intended in the second person, non-honorific, singular, the personal ending =*m* '2sg' is omitted, as in (96)–(98). If a future or hypothetical situation is being narrated, the subject marker is indicated, as in (99). This applies only to the second person, singular (non-honorific). The second persons dual/honorific and plural are always marked for PERS/NUM/HON-status, as in (95).

93. *tobdaʔ daʔ=te i=ghay uq=e=niŋʔ*

mud water=OBL what=way drink=ACT.IRR=1PL.INCL

'How *will we drink* muddy water?' [MS, 1:126]

94. *muda dada=dom=ki=yaʔ aw-aw=te ber*
but elder.brother=3POSS=PL=GEN QUAL-RDP=OBL who

beʔi=ki=te um uqun=na pal=e.

dather=PL=OBL NEG make.flee=INF be.able=ACT.IRR

'But as long as the elder brothers are present, no one *will be able to drive away* the daughters.'
[MT, 1:169]

95. *kulu gam=oʔ "baru kayom! iŋ=te=jo am=bar=aʔ*
turtle say=ACT.PST good speech 1SG=OBL=ADD 2=2DU=GEN

sori dɔʔ=e=bar!"

with take=ACT.IRR=2DU

'The turtle said "Great idea! *Take* (DU) me along with you as well!"'
[TK, 1:11]

96. *ho=kaṭ kontheḍ=te gam=oḥ, “gerwa bohin,*
 that=SG.HUM bird=OBL say=ACT.PST nightingale sister
a=e iṅ=a[ʔ] sori kuʔj=na, eboʔ=na.”
 come.ON=ACT.IRR 1SG=GEN with dance=MID.IRR play=MID.IRR
 ‘He said to the bird, “Sister nightingale, *come on! Dance and play*
with me!”’ [TK, 2:12]

When used to denote an imperative, an “emphatic particle” such as *hen* ‘come on!’ in the following example may also be used:

97. *etwa dular buṅ gam=na laʔ=ki “hen beḥa=iṅ,*
 Etwa love INST say=INF IPFV=MID.PST come.on! son=1sg
kaṭiʔj-ḍuʔ uʔḍ goṭ=e.”
 a.little-APPROX drink C.TEL=ACT.IRR
 ‘Etwa would say lovingly “*Come on, my son! Drink a little bit!*”’
 [RD, 2:8]

After vowels, the irrealis active is realized as =ye:

98. *yo=ye ketna sundar mawsam ayiʔj.*
 see=ACT.IRR how.much beautiful weather QUAL.PRS
 ‘*Look* how beautiful the weather is!’ [TK, 2:13]

The irrealis active marker =e is omitted following morphemes ending in -y and preceding the personal endings =iṅ ‘1sg’ and =em ‘2sg’. For ease of comprehension, in these environments this will be indicated in texts as “[=e]”, as in the following examples. It is important to stress here that this does not indicate that the text is corrupt, as this is a productive phonotactic rule of the language (cf. 2.4):

99. *musa tay am puḅapaḥ karay[=e]=em.*
 today ABL 2SG sacrifice do=ACT.IRR=2SG
 ‘As of today, you *will perform* sacrifices.’ [AK, 2:33]
100. *u lereʔ buṅ iṅ am=pe=te tuḍa raylo kimir*
 this joy INST 1SG 2=2PL=OBL tomorrow Railogarh forest
hinte lam-lam co=na=ʔ thoṅ chuṭi ter[=e]=iṅ.
 LOC search-RDP go=INF=GEN PURP permission give=ACT.IRR=1sg
 ‘Because of this joy I *will give* you permission to go hunting in
 Railogarh Forest tomorrow.’ [MT, 1:91]

With the middle irrealis marker =*na*, the /i/ in the marker of the first person, singular, =*ijn*, may also be omitted, although this is optional, even for one and the same speaker/writer, as in the following two examples:

101. *am=bar lej=na=bar la? ijt=jo laṭke=ga*
 2=2DU fly=MID.IRR=2DU then 1SG=ADD swing=FOC
 co=*na=ijn*
 go=MID.IRR=1SG
 'You fly and I *will* also go swinging.' [TK, 1:19]

102. *kulu cilay=ki, "ijn um=ijn kayom=na? jarur*
 turtle shout=MID.PST 1SG NEG=1SG speak=MID.IRR certainly
kayom=na=ṅ!"
 speak=MID.IRR=1SG
 'The turtle shouted, "I shouldn't speak? Of course I *will speak!*"' [TK, 1:36]

The marker of the second-person, singular, on the other hand, always has the form =*m* after =*na*, yielding =*na=m*:

103. *e bhai, am kayom=na=m la? dāṭa? i=ghay*
 voc brother 2SG speak=MID.IRR=2SG then stick what=way
dhog[=e]=em?
 grab=ACT.IRR=2SG
 'Oh brother, if you *speak*, then how will you hold on to (= grab) the stick?' [TK, 1:37]

The non-imperative meaning of the irrealis is negated by the general negator *um* (6.7.1) whereas the imperative meaning is negated by *abu* (6.7.2).

6.4.1.5 Perfect

Present perfect The present perfect is marked by the perfect marker *siṭḍ*. In its primary function, it denotes that an event has occurred which is of direct relevance to a later situation, usually the speech act.

With reference to a single, particular situation, the present perfect is unmarked for TAM/basic voice. For example, in the following example, a grasshopper has neglected to gather food for the cold season and has nothing to eat when he suddenly remembers that all other animals

have gathered food. The TAM/PERSON-syntagma appears in the perfect as the fact that the others did gather food is of direct relevance to what follows, namely that the grasshopper will now go ask the others for food.

104. *ho=kaṛ socay=oʔ, “sob ranḡa bhère=yaʔ ḡhaḡ*
 that=SG.HUM think=ACT.PST all cold time=GEN for
noʔ=na thoṅ thuray=siʔ=may. col=ta=ṅ ho=ki=yaʔ
 eat=INF PURP gather=PERF=3PL go=Mid.PRS=1SG that=PL=GEN
tay noʔ=na thoṅ bor ol[=e]=iṅ.”
 ABL eat=INF PURP ask.for bring=ACT.IRR=1SG
 ‘He thought, “All *have gathered* [food] for eating for the cold season. I will go beg for something (= ask for [and] bring) to eat from them.”’
 [TK, 2:31–32]

The second major function of the perfect is what is often termed the “experiential perfect”, i.e. to denote that some event / state is depicted as being within the subject’s sphere of experience:

105. Q: *am=bar “Rock Garden” deʔb=siʔ=barʔ*
 2=2HON Rock Garden ascend=PERF=2HON
 A: *hã deʔb=siʔḡ=iṅ.*
 yes ascend=PERF=1SG
 Q: ‘*Have you ever climbed* Rock Garden (the name of a park on a hill in Ranchi)?’
 A: ‘Yes, *I have climbed* it.’

This question was posed at a time when neither of the two speakers was located on Rock Garden hill, thus the only possible interpretation of the perfect is whether the addressee has ever performed the action.

The present perfect, when followed directly by the conjunction *laʔ* ‘then’, is also used to denote a future action which is directly followed by another action. Consider the following example:

106. *raksin oḡoʔ juḡhay khisay=ta ro gone keb=kon*
 witch even more angry=Mid.PRS and tooth grind=SEQ
gam=te: “dhoʔ=siʔḡ=iṅ laʔ am=aʔ
 say=ACT.PRS grab=PERF=1SG then 2SG=GEN

haṛ=no=m=te=jo eb=kon thom=ta thomta akeḍ
 bone=2POSS=2SG=OBL=ADD roast=SEQ pulverize=CVB REP chew

akeḍ jog[=e]=iṅ."

REP eat=ACT.IRR=1SG

'The witch grows even angrier and, grinding her teeth, says "I will grab you and then, having roasted your bones too, I will eat them, chewing and thereby pulverizing them."' [BB, 1:49]

As noted in Section 2.4, the perfect morpheme is optionally realized as =*chi(?)ḍ* following a morpheme-final /j/:

107. *kongher rag badli=kon gam=oḥ "gaʔj=chiḥiṅ,*
 boy voice change=SEQ say=ACT.PST fry=PERF=1SG

yo, ol kay=t[e]=iṅ?"

mother.VOC bring BEN=ACT.PRS=1SG

'The boy, changing his voice, said "I have fried [the meat], mother, shall I bring some for you?"' [BB, 1:80]

In all of the examples so far, the present perfect is a "present" category by default, as there is no marking for the present tense. In fact, this only holds for the present perfect when it refers to a SINGLE, PARTICULAR SITUATION. When a habitual situation is referred to, =*siḥiḍ* combines with either the active or middle present-tense markers:

108. *jughay=aḥ=ko umboḥ moṅ patha=? bhāḥa=ko*
 much=GEN=CNTR NEG one.NUM kilo=GEN pot=CNTR

un=siḥ=te=ga jahāy awal gel gam=kon.

place=PERF=ACT.PRS=FOC INDEF "guest" say=SEQ

'She *doesn't* [usually] place [rice beer] made from (= of) too much [rice], only a pot from one kilogram for guests (= having said "guest").' [Kerkettā, 1990: 27]

109. Q: *akhaṛ kinir=te moṅ bacha tol ḍom=siḥ=ta.*
 deep forest=OBL one calf tie PASS=PERF=MID.PRS

A: *kusar* [HJPa:184, ln. 127]

cocoon.of.silk.worm

Q: [What is this?] A calf is [regularly] tied in the deep forest.

A: The cocoon of the silk-worm.' (a riddle, referring to something that is habitually found tied in the forest).'

Cf. also HPA:176, ln. 60 for two further examples in the active.

The perfect is also compatible with the markers for the simple past tense and for the irrealis, although apparently not with the progressive.¹⁵ The past and irrealis perfect are discussed in the following two sections.

Past perfect The past perfect is found only in the active. It usually designates a past event which preceded another event in the past.

110. *am=pe, iŋ gam=sikh=oʃj ho=ghay=ga am=pe*
 2=2PL 1SG say=PERF=ACT.PST.1SG that=way=FOC 2=2PL
col=ki=pe.
 go=MID.PST=2PL
 ‘You, you went just as I *had told* you.’ [AK, 1:63]

111. *pheinga gerwa konthed=yaʔ boʔ=te col=ki*
 grasshopper nightingale bird=GEN place=OBL go=MID.PST
muda gerwa boŋ bheʔto um hoy=ki.
 but nightingale INST meeting NEG become=MID.PST
ho=kaʔ garam raij tiʃ col kan=sikh=oʔ.
 that=SG.HUM warm kingdom side go CONT=PERF=ACT.PST
 ‘The grasshopper went to the house of the nightingale but he did not meet the nightingale (= a meeting with the nightingale did not become). She *had gone off* to a warm place.’ [TK, 2:36–37]

The past perfect may also be used as a simple past tense. From this usage, the Past II discussed in Section 6.4.1.1 has developed.

112. *ho=kaʔ ho jaraʔ daru=yaʔ khoʔri=te murti=te*
 that=SG.HUM that banyan tree=GEN hollow=OBL statue=OBL
ko<?>sor=na un=sikh=oʔ.
 dry-<CAUS>=INF place=PERF=ACT.PST
 ‘He *placed* the statues in the hollow of that banyan tree to dry [them].’ [AK, 3:13]

¹⁵ That is, I do not have any examples for this in my own corpus and have not found any in the first 143 pages of Pinnow (1965a), although this could be due to coincidence. At any rate, even if it is compatible with the progressive, its use is extremely seldom.

Although the past perfect is restricted in the modern language to the active, at least that of all speakers we worked with and in all modern texts we have analyzed, there are nevertheless at least two examples of the past perfect occurring in the middle voice in Pinnow (1965a), both involving a change of state:

113. loʔd̪ho=siʔ=ki=may ho=ki “poʔoŋ poʔa!” gam=oʔ ro
 after=PERF=MID.PST=PL that=PL hare stomach say=ACT.PST and
 noʔ goʔh=oʔ=ki.
 eat C.TEL=ACT.PST=PL
 ‘Those who had followed (= *had become later*) said “A hare’s stomach!” and ate it. [HJPa:124, ln. 85]
114. ma=d̪om khisay=siʔ=ki, magar i=ye, girja
 mother=3POSS angry=PERF=MID.PST but what=ACT.IRR church
 col kan=ki kund̪uʔ=ki=te d̪oʔd̪=k̪on.
 go CONT=MID.PST child=PL=OBL take=SEQ
 ‘Their mother *had become angry*, but what should she do (= she should what?), she went off to church with (= having taken) the children.’ [HJPa:57]

Interestingly, speakers questioned on these sentences rejected both as ungrammatical. As these seem to be unique examples (although they occur in different texts in Pinnow’s study) it would seem best to consider them idiolectal forms, although it is also possible that there once was a productive opposition in the past perfect between the active and middle which has since been lost. Perhaps future work on other Kharia dialects will turn up similar forms.

Irrealis perfect Unlike the past perfect, which is restricted to the active in the modern language, the irrealis perfect is found in both the active and the middle. It is found in the same environments as the irrealis, i.e., future events, imperatives and conditional clauses.

115. oʔ=te ap=d̪om teinko arkhi un=sig=e
 house=OBL father=3POSS a.little liquor place=PERF=ACT.IRR
 ho=jeʔ=ko ho=kaʔ capu karbaʔd̪=k̪on uʔd̪
 that=SG.NHUM=CNTR that=SG.HUM rummage ECHO=SEQ drink

iɖaʔ=na *laʔ=ki.*

ECHO=INF

IPFV=MID.PST

‘If his father *has placed* some liquor in the house, then he would rummage through everything [until he found it] and then drink it all up.’ [RD, 2:25]

116. *koroʔb=siʔ=na=pe.* *ber=jo* *i=jo* *aʔ=pe*
 silent=PERF=MID.IRR=2PL who=ADD what=ADD NEG.MOD=2PL

gam=e.

say=ACT.IRR

‘*Be quiet!* Don’t any of you say anything.’ [Kerkeṯṯā, 1990: 2]

Finally, the perfect is also compatible with infinitival marking:

117. *ho=tiʔj* *modi* *ɖay=jo* *koleʔj=na=ʔ* *thoŋ*
 that=side Modi woman=ADD fight=INF=GEN PURP

aŋ=siʔ=na *laʔ=ki.*

open.mouth=PERF=INF IPFV=MID.PST

‘There Modi’s wife would also be fighting (= *used to open [her] mouth* for fighting).’ [RD, 2:73]

Perfect vs. non-perfect As the following examples show, the perfect is often found in environments in which we would expect the present from an English-speaking perspective. This is the case with change-of-state predicates (known by a variety of different names, including “inchoatives”, “inceptives”, “transformatives”, etc.):

118. *tuyu* *beroɖ=ki* *ro* *soloʔ=te* *gam=oʔ* *no* “*am=te*
 jackal arise=MID.PST and dog=OBL say=ACT.PST Cmpl 2sg=OBL

lemeɖ *um* *ɖel=ta* *laʔ* *korob* *korob*

sleep NEG come=MID.PRS then silent REP

gitaʔ=siʔ=na, *umboʔ* *laʔ=ko* *ɖoko=siʔ=na.”*

lie.down=PERF=MID.IRR NEG then=CNTR sit.down=PERF=MID.IRR

‘The jackal got up and said to the dog “If you can’t sleep then *lie* quietly, otherwise *sit* [quietly].”’ [BB, 3:39]

119. *in dāra?=te moʔjhi=te gone buj akeʔd=sig[=e]=in.*
 1SG stick=OBL middle=OBL tooth INST bite=PERF=ACT.IRR=1SG
 'I'll hold on to (= *will have bitten*) the stick with my teeth.'
 [TK, 1:18]
120. *hāyre arabduʔ moʔn=ga bohini=na=in daru sumboʔ=te*
 alas poor one=FOC sister=1POSS=1SG tree base=OBL
japaʔ=siʔ.
 lean.against=PERF
 'Alas, my poor sister *is leaning* (= *has leaned*) against a tree.'
 [BB, 4:2]

The fact that the perfect is found in such examples does not mean that it is required by the respective contentive morpheme. Cf. e.g. the following examples with those above:

121. *“korob korob gitaʔ=na.” soloʔ=te kirog=aʔ*
 silent REP lie.down=MID.IRR dog=OBL tiger=GEN
kayom baru um laʔ=ki.
 speech good NEG EMOT=MID.PST
 ‘“*Lie down quietly.*” The dog did not like the tiger’s words.’
 [BB, 3:54]
122. *betoʔd buj modi=yaʔ konon kulam kulam-day=dōm=ki*
 hunger INST Modi=GEN small sibling sibling-woman=3POSS=PL
bhit sinij sinij japaʔ khor=taʔj=may bhere moʔd kici
 wall edge REP lean ITER=MID.PROG=3PL time eye sleep
oqoʔ rajh rajh romoʔddaʔ gur=taʔj.
 and falling.in.quick.procession tear fall=MID.PROG
 ‘When Modi’s younger brothers and sisters *were leaning*¹⁶ here and there (= *chor*) against the wall in hunger, their eyes [were full of] sleep and their tears were falling in quick procession.’ [RD, 2:56]

¹⁶ This is a case of the ‘historical present’ use of the progressive, see Section 6.4.1.3.

It appears that the use of the perfect in examples such as these with change-of-state predicates emphasizes the state which is the result of this change, and not merely the action itself, e.g. 'lie down [and be lying]' vs. 'lie down', 'lean against [and be leaning]' vs. 'lean against'

si? vs. *siʔd* Recall from Section 2.4 that a morpheme-final /ʔ/ (< */g/) is realized as /kh/ before the past active marker =*oʔ*. Considering this, the fact that the past perfect has the form *sikh=oʔ* suggests that the underlying form is *si?* (< /sig/) and not *siʔd*. Similarly, the active irrealis perfect in =*sig=e* seems to point in this direction as well. Although the original form of this morpheme was apparently *siʔd*, as it has cognates in other Munda languages (Greg Anderson, p.c.), from a purely synchronic perspective the case is not quite as clear.

In the present perfect, this marker has the form /*siʔd*/ before =*ɨn* and =*em*, the markers of the first and second persons, singular, respectively, whereas the form is realized as /*si?*/ elsewhere. These are also the only two personal markers which begin with a vowel. The same distribution of /*d*/ is also found in the same two persons with the non-inherent qualitative predicative marker *ayiʔj* (7.3), i.e., *ayiʔjd* vs. *ayiʔj*, as well as with the present progressive markers =*teʔjd* / =*teʔj* (ACTIVE) and =*taʔjd* / *taʔj* (MIDDLE) (6.4.1.3). In the case of the qualitative predicative marker and the progressive markers, it was argued in Section 2.4 that the deletion of /*d*/ is due to the strict phonotactic rule that neither complex onsets nor codas are allowed, i.e., **ayiʔjd=le* 'we (EXCL) are' > *ayiʔj=le*. This explanation, however, will not apply in the case of the perfect, unless we consider the /ʔ/ to be a consonant in *siʔd*, and not merely the pre-glottalization of a morpheme-final /*d*/ . This analysis is strongly supported by the fact that when the /*d*/ is not present, as in *col=si?* 's/he has gone', the glottal stop is nevertheless present, i.e., it cannot be mere pre-glottalization as there is no pre-glottalized consonant. The perfect marker is therefore analyzed here as having the underlying form *siʔd*, bringing it in line with our analysis of the qualitative predicative marker and the progressive markers.

On the other hand, it is possible to consider the underlying form of the perfect to be *si?*, with /*d*/ added before the markers of the first and second persons, singular. It is undoubtedly for this reason that Biligiri (1965: 59) speaks here of an "increment /*d*/" for both the perfect and the progressive (or "present immediate" in his terminology), an analysis that

is transferable to the qualitative predicative marker as well.¹⁷ This has the advantage of being able to account for both the past and irrealis perfect more elegantly than our analysis, which however has the advantage of offering a uniform analysis of the alternating forms of the present perfect, the qualitative predicative marker and the progressive markers.

The historical development seems clear: It is quite common in Kharia for pre-glottalized consonants in the coda to be realized in speech simply as the glottal stop, e.g., one hears *kaʔiʔ* just as often as the "standard" form *kaʔiʔj* 'a little' or *goʔjhuŋ* equally as often as *goʔdʒhuŋ* 'path, way'. In fact, some speakers are not even familiar with the "standard" forms of such morphemes, and the original pre-glottalized consonant has developed into a simple glottal stop for these contentive morphemes for these speakers.

This is obviously what has happened with the original perfect marker *siʔd*, which has now become *siʔ* in all environments except for the first and second persons, singular in the present perfect. Thus, this marker may either be analyzed as *siʔ*, with an increment /d/ exceptionally added in these two persons along the lines of the alternation between /d/ and Ø found in the qualitative predicative marker and the progressive, or it has the underlying form *siʔd*, and the /d/ is regularly deleted before personal markers beginning with a consonant in the present perfect, and irregularly in both the past and irrealis perfects.

Either analysis seems equally justified and both have exceptions. This form is analyzed here as =*siʔd* mainly for convenience of glossing and to avoid having to incorporate an "increment" which is found nowhere else in the language.

6.4.1.6 *Optative*

The optative is used to denote a strong wish or obligation in the third person or for third-person imperatives. It may not be used in the first or second persons.

Unlike all other basic TAM categories, *guʔuʔ* / *guʔuʔ* is a phonological word and may be separated from preceding elements of the TAM/PERSON-syntagma, such as the semantic base or a v2 (6.5), by the focal particle =*ga*. However, it has no independent lexical meaning:

¹⁷ Biligiri (1965: 64) however analyzes the qualitative predicate marker as *ai* with the progressive marker *-j* (termed "present immediate" on page 58 in his grammar), an analysis which I reject since the progressive would then be compatible only with the present but with no other TAM-categories, although this is admittedly a rather arbitrary decision.

123. *maha bhai mana karay=si? tewa?jo aḍi jarur*
 big brother forbid do=PERF nevertheless ANAPH certainly
ḍel=ga guṛu?
 come=FOC OPT
 ‘Although his_i elder brother has forbidden it, nevertheless he_i *should*
certainly come.’

124. *aniṅ=a? konsel kongher=ki khaṛiya=ki=ya? bair*
 1PL.INCL=GEN girl boy=PL Kharia=PL=GEN old
nāḍani=te koṅ guṛu?=may ro khaṛiya=ki=ya? paṛom
 history=OBL know OPT=3PL and Kharia=PL=GEN strength
ro main=te koṅ=kon oḍo? meson ho=je? arjay
 and honor=OBL know=SEQ more once that=SG.NHUM revive
guṛu?=may.
 OPT=3PL
 ‘Our girls and boys *should know* the old history of the Kharia and,
 having learned the honor and strength of the Kharia, they *should*
revive it once again.’ [Kerkeṭṭā, 1990: ii]

In some cases, the use of the optative closely resembles that of a subjunctive. Here as well, however, it is restricted to third-persons:

125. *paṛom ro iḅat am=pe=ga. i, am=pe lam=te-pe*
 energy and respect 2=2PL=FOC Q 2=2PL want=ACT.PRS=2PL
no kesariya=ki ḍel=kon raij=te raijhay guṛu?=may?
 Cmpl Kesari=PL come=SEQ kingdom=OBL rule OPT=3PL
 ‘You are [this country’s] energy and honour. Do you want the Kesari to come and rule the kingdom (= that the Kesari, having come, *should rule*)?’ [Kerkeṭṭā, 1990: 5]

The optative is negated by the marker *abu* (6.7.2), similar to the imperative use of the irrealis.

6.4.2 Active and middle: Their functions and distribution

6.4.2.1 General remarks

The first four basic TAM categories in Table 6.15 above are expressed by portmanteau morphemes which simultaneously express basic voice, i.e.,

active and middle. These four categories are repeated here as Table 6.16 for ease of reference. In 6.4.1 the TAM values of these morphemes were discussed in detail. The present section discusses the functions of the active and middle. These categories are obligatory: A finite TAM/PERSON-syntagma appearing in any one of the TAM categories given in Table 6.16 must always be marked as either active or middle.

Table 6.16: TAM/basic voice markers

	<i>Middle</i>	<i>Active</i>
Past (PST) (aspectually neutral)	= <i>ki</i>	= <i>(y)o?</i>
Present General Imperfective (PRS)	= <i>ta</i>	= <i>te</i>
Present Progressive (PROG)	= <i>ta'jq</i>	= <i>te'jq</i>
Irrealis (IRR)	= <i>na</i>	= <i>e</i>

Before beginning our discussion, a few words are in order on the choice of the terms “active” and “middle”: Note that it is usual to speak of these two categories in Kharia—and Munda languages in general—not as basic voice but rather as “transitive” (ACTIVE) and “intransitive” (MIDDLE) or to merely divide them into “Set 1” (ACTIVE) and “Set 2” (MIDDLE), whereby a more or less direct link to transitivity is usually assumed.

Although these two classes are closely connected to (in-)transitivity, as we shall see in the following sections, they are also productively used to express a number of distinctions which are not directly related to (in-)transitivity. Rather, a comparison with the active and middle voices in other languages from different families¹⁸ shows that, despite all differences, the two categories in Kharia share many traits with these systems.

Following Klaiman (1991), these categories will be referred to collectively as basic voice, as neither is more marked than the other, morphologically speaking, and one of the two is obligatorily present in the four basic TAM categories. This is in contrast to derived voice, such as the passive/reflexive (6.5.1), which contrasts with zero and is clearly a marked voice.

Again following Klaiman (1991) the functions of the active and middle are divided into into their “differential” and their “inherent” functions: Differential functions (6.4.2.2) are those functions which are relevant for contentive morphemes which can appear in both the active and the middle

¹⁸ E.g., in chronological order, Gonda (1979), Klaiman (1991), Kemmer (1993), Kaufmann (2004) and Brill (2005).

voice while inherent functions (6.4.2.3) are those “functions” which are relevant for contentive morphemes which appear in either the active or the middle, but not in both. These inherent “functions” are, properly speaking, not functions but merely tendencies, but this term will be retained here for the sake of consistency.

As we shall see, the inherent functions are in fact merely a sub-set of the differential functions, and contentive morphemes which can only appear in one of these two categories but not in both receive their voice assignment depending on whether their basic usage shows more similarities to the differential functions of the middle or to those of the active.

6.4.2.2 *Differential functions*

While the large majority of contentive morphemes in Kharia which are compatible with both the middle and active have the same form in both categories, it was shown in 6.3.3 that there is a considerable number of contentive morphemes which have been borrowed from Sadri which have two different forms, generally ending in *-e* in the middle and *-ay* in the active. Tables 6.17 and 6.18 present a few examples of both types.

Table 6.17: Morphemes with one form in both middle and active

	<i>Middle</i>	<i>Active</i>
<i>bel</i>	‘spread out (of people)’ (ITR)	‘spread out (of people, cloth, etc.)’ (TR)
<i>ga?</i>	‘rip (of paper)’ (ITR)	‘rip (of paper)’ (TR)
<i>pa?</i>	‘break’ (ITR)	‘break’ (TR)
<i>ro?</i>	‘spill’ (ITR)	‘spill’ (TR)
<i>sango’d</i>	‘close’ (ITR)	‘close’ (TR)

Table 6.18: Morphemes with different forms in middle and active

	<i>Middle</i>		<i>Active</i>
<i>bagre</i>	‘spoil (ITR)’	<i>bagray</i>	‘spoil (TR), ruin’
<i>cale</i>	‘function; go’	<i>calay</i>	‘drive (a car, etc.)’
<i>chu?te</i>	‘leave’ (ITR)	<i>chu?tay</i>	‘leave’ (TR)
<i>khole</i>	‘open’ (ITR)	<i>kholay</i>	‘open’ (TR)
<i>salge</i>	‘burn, catch fire’	<i>salgay</i>	‘light a fire’
<i>thane</i>	‘try unsuccessfully to assert’	<i>thanay</i>	‘assert’

A. Middle: Intransitive, Active: Transitive As the data in Tables 6.17 and 6.18 show, the middle and active are intimately connected to (semantic) transitivity.¹⁹ In these cases, the middle is regularly intransitive and the active transitive. Table 6.19 presents a few further examples.

Table 6.19: Transitivity and basic voice

	<i>Middle</i>	<i>Active</i>
<i>ayo</i>	'become a mother'	'accept [s.o.] as a mother'
<i>ayo=ya?</i>	'become mother's'	'make [s.th.] mother's'
mother ^{I=GEN}		
<i>iji=a?</i>	'become mine'	'make [s.th.] mine; adopt [s.o.]'
1 ^{SG=GEN}		
<i>kuṭum</i>	'become family (e.g through marriage)'	'accept [s.o.] as family'
<i>khatam</i>	'come to an end'	'finish' (TR)
<i>saŋgo'd</i>	'close' (ITR)	'close' (TR)
<i>se?</i>	'fill up' (ITR)	'fill up' (TR)
<i>sebol</i>	'become sweet'	'make sweet, sweeten'
<i>tuta</i>	'go down'	'put down, lower'
<i>thorek</i>	'decrease' (ITR)	'decrease' (TR)

The members of this group denote events which may either be caused by an external agent (the subject in the active) or may simply occur on their own, in which case the patient is the subject (in the middle voice). This feature distinguishes them from transitive active-only contentive morphemes and intransitive middle-only contentive morphemes: In active-only transitives, the action is conceived of as requiring an external agent (e.g., *go?* 'carry' or *ji'b* 'touch') whereas middle-only intransitives are conceived of as denoting an event which generally occurs on its own (e.g., *go?j* 'die'). If an external agent is to be expressed here, either a different contentive morpheme is used (e.g., *tar* 'hit, beat, kill') or the TAM/PERSON-syntagma appears in the causative (e.g., *o²b-go?j* 'cause to die').

This does not mean that the classification of morphemes into these three groups according to natural propensity can be deduced intuitively. Rather, the language has lexicalized this distinction in a particular way

¹⁹ As we shall see in 7.1, syntactically speaking all predicates in Kharia are intransitive, requiring only the subject marking on the predicate. All Case-syntagmas are thus "adjuncts", as their presence is never required. I will thus use the term "transitivity" here in a semantic sense, i.e., those elements which are "obligatory" at the semantic level, not at the syntactic level.

for each contentive morpheme. Thus, determining whether a contentive morpheme is classified as denoting an event which typically happens on its own, is typically caused by an external agent, or is compatible with either interpretation, is a purely empirical question.

There is considerable variation in the linguistic literature concerning the designation of these transitive/intransitive pairs. In the more theoretically oriented literature, they are often termed “ergative-causative pairs” (cf. e.g. the discussion in Haegeman, 1991: 334–336), while other authors consider the intransitive members to be a sub-group of the so-called “unaccusatives”, such as Levin & Rappaport Hovav (1995), who speak of causatives and (unaccusative) non-causatives in the “causative alternation” Klaiman (1991) on the other hand refers to the intransitive member of such pairs as “neuters” These will simply be referred to here as “transitive/intransitive” pairs.

Before continuing, it will be instructive to consider the Kharia data with respect to transitivity with the data discussed in Nichols et al. (2004), who deal with “transitivizing and detransitivizing languages” from a cross-linguistic perspective. Briefly summarized, Nichols and her associates refer to the semantically non-causative form as the “plain” form and the semantically causative form as “induced” The authors then undertake a study of the means of deriving one of these forms from the other in a number of languages, i.e., is a particular language primarily a transitivizing or detransitivizing language? As we shall see, Kharia is primarily a transitivizing language.

The three means of arriving at the induced form which are relevant for our discussion of Kharia are augmentation, which in Kharia can be understood as referring to the causative marker //OʔB//, as it is added to the plain form, suppletion, and “class change”, which in our context can be equated with the Middle / Active opposition in Kharia.²⁰

Although many contentive morphemes have a transitive sense when they appear in the active voice as opposed to an intransitive sense in the middle voice, the active voice does not appear to be the primary means of denoting transitivity in Kharia, at least not according to the plain/induced pairs in the sample in Nichols et al. (2004). Table 6.20 shows the conten-

²⁰ With respect to conjugation class change, the authors write “. . . conjugation classes are inflectional classification rather than derivational categories, verbs showing this correspondence belong to different conjugation classes in their intransitive and transitive uses but are otherwise underived” (Nichols et al., 2004: 159).

tive morphemes listed in Nichols et al. (2004) and their respective means of transitivization in Kharia.

Table 6.20: "Transitivizing strategies" in Kharia (based on Nichols et al., 2004)

<i>Pair</i>	<i>Plain</i>	(* = active, otherwise middle)	<i>Induced</i>	(all in active)	
1.	'laugh'	<i>*laḍa</i>	'make laugh'	<i>la<ʔb>ḍa</i>	Augmented
2.	'die'	<i>goʔj</i>	'kill'	<i>oʔb-goʔj</i>	Augmented, class change;
				<i>tar</i>	Suppletion
3.	'sit'	<i>ḍoko</i>	'seat'	<i>ḍo<ʔb>ko</i>	Augmented, class change
4.	'eat'	<i>*jioʔ</i>	'feed'	<i>ob-jioʔ</i>	Augmented
5.	'learn'	<i>sike</i>	'teach'	<i>sikhay</i>	Augmented, class change
6.	'see'	<i>*yo</i>	'show'	<i>oʔb-yo</i>	Augmented
7.	'become angry'	<i>khisay</i>	'anger'	<i>oʔb-khisay;</i> <i>khisway</i>	Augmented, class change
8.	'fear, be afraid'	<i>botoŋ</i>	'frighten, scare'	<i>bo<ʔʔ>toŋ</i>	Augmented, class change
9.	'hide'	<i>leʔḍ</i>	'hide'	<i>oʔb-leʔḍ</i>	Augmented, class change
10.	'boil'	<i>uluʔ</i>	'boil'	<i>u<ʔʔ>luʔ</i>	Augmented, class change
11.	'burn, catch fire'	<i>geʔb</i> <i>salge</i>	'burn, set fire'	<i>o-geʔb</i> <i>salgay</i>	Augmented, class change
12.	'break'	<i>paʔ; piʔj</i>	'break'	<i>paʔ; piʔj</i>	Class change
13.	'open'	<i>ruʔ</i>	'open'	<i>ruʔ</i>	Class change
14.	'dry'	<i>kosor</i>	'make dry'	<i>ko<ʔb>sor</i>	Augmented, class change
15.	'become straight'	<i>seŋghor</i>	'straighten'	<i>seŋghor</i>	Class change
16.	'hang'	<i>ṭaŋe</i>	'hang'	<i>ṭaŋay</i>	Augmented, class change
17.	'turn over'	<i>kulaŋ;</i> <i>ulṭe</i>	'turn over'	<i>ku<ʔb>laŋ;</i> <i>ulṭay</i>	Augmented, class change
18.	'fall'	<i>gur</i>	'drop, let fall'	<i>o-gur</i>	Augmented, class change

Note that pairs 1–9 "have varying degrees of agency and volition on the part of an animate S/O; 10–18 have varying degrees of independence, resistance to force, etc. on the part of an inanimate S/O." (Nichols et al.,

2004: 155–6). It is significant that all of the first nine contentive morphemes require augmentation to derive the induced form from the plain form, and class change, i.e., active vs. middle, is secondary to this. On the other hand, class change alone is the sole means of transitivization for some forms in 10–18, although even here it is not very common (3 out of 9 forms).

To the extent that these contentive morphemes are indeed representative of the basic transitivizing strategy in Kharia, we conclude that the primary transitivizing strategy in Kharia is augmentation by means of the causative, whereas class change alone is quite limited, although more common than suppletion. More work is required on this topic, especially to determine the actual frequency of augmentation by means of the causative (with secondary class change) vs. class change alone, as the three forms here which signal transitivity solely through the active (vs. middle) voice are in fact members of a much larger group in Kharia.

Although (in-)transitivity is undoubtedly the predominant differential (and inherent) function of the middle and active, it is only one of a number of more or less productive differential functions. These other functions are dealt with in the following sections.

B. Middle voice as “generic” The generic middle function is the only completely productive differential function of the language and applies whether the underlying contentive morpheme is otherwise restricted to the middle, the active, or whether it can appear in both categories. In this function, all TAM/PERSON-syntagmas appear in the middle voice only.

In this function of the middle, semantic bases of TAM/PERSON-syntagmas must be the masdar (6.6.2.1): Monosyllabic contentive morphemes reduplicate, whereas polysyllabic (derived or underived) forms (= contentive morpheme and, if present, causative / passive marking and non-telic v2s, cf. 6.5) remain unchanged. E.g. *eŋ* ‘return (TR)’, masdar: *eŋ-eŋ*; *oyeŋ* ‘return (TR)’, causative of *eŋ*, masdar *oyeŋ*, *rusuŋ* ‘(become) red’, masdar *rusuŋ*. Table 6.21 gives a few examples of morphemes appearing in their unmarked usage (here with examples of otherwise middle-only and active-only contentive morphemes) and the corresponding form of the masdar as the semantic base in the generic function.

Referring to the non-generic functions of these morphemes for convenience as “neutral”, we may describe the function of the generic middle as follows: Whereas the neutral function is unmarked, the generic function expresses habituality or long-term iterativity with a reduplicated masdar.

Table 6.21: Middle as "generic"

<i>Middle-only</i>	<i>Active-only</i>	<i>Generic Middle</i>
	<i>absiʔb</i> 'begin'	<i>absiʔb</i>
	<i>biʔd</i> 'pour out; sow'	<i>biʔd-biʔd</i>
	<i>buʔ</i> 'beat (a drum)'	<i>buʔ-buʔ</i>
<i>boŋ</i> 'finish'		<i>boŋ-boŋ</i>
<i>eŋ</i> 'return'		<i>eŋ-eŋ</i>
	<i>hakɾe</i> 'roar, groan (of animals)'	<i>hakɾe</i>
	<i>koʔpeʔd</i> 'pull together'	<i>koʔpeʔd</i>

Let us begin by way of example with a simple contentive morpheme as the semantic base in predicative function, *biʔd* 'pour out'. This morpheme, when used predicatively, is restricted to the active voice and denotes, e.g., in the past tense that the subject poured something out but does not further specify whether this was a one-time action or whether it was performed more than once. When the masdar of this morpheme is used as the semantic base and the TAM/PERSON-syntagma is marked for the middle voice, on the other hand, it is a habitual action, e.g., when it is part of someone's job to pour water out of something over and over:

126. *ɪŋ daʔ biʔh=ɔʔj.* *Unmarked construction*
 1sg water pour.out=ACT.PST.1sg
 'I poured water out.'
127. *ɪŋ daʔ biʔd-biʔd=ki=ɲ.* *Generic middle*
 1sg water pour.out-RDP=MID.PST=1sg
 'I used to pour water out. (e.g., that was my job, so I did it all the time)'

The reduplication of the underlying contentive morpheme here, in conjunction with the middle voice, is iconic to a certain extent in that the habitual reading is largely restricted to masdars formed by the reduplication of an underlying monosyllabic morpheme.

As noted in 4.3.2, the use of the non-reduplicated masdar as the semantic base in the generic middle is one of a number of "shibboleths" differentiating the Simdega dialect from the Gumla dialect: While speakers of the Simdega dialect accepted this construction with masdars of both mono- and polysyllabic contentive morphemes, speakers of the Gumla dialect rejected such non-reduplicated masdars in the generic middle. Thus, at least based on the data collected from interviews, it would seem

that the generic middle is only completely productive in the Simdega dialect whereas in areas further to the north and east it is restricted to the expression of habituality/iterativity with reduplicated monosyllabic contentive morphemes as the semantic base. The following comments thus apply only to the Simdega dialect.

With the non-reduplicated masdar as the semantic base, the generic middle can have a habitual interpretation, however, this is not its default interpretation. Instead, the default interpretation is one of the following, depending on tense marking and context:

- The generic middle denotes that an event took considerably longer than usual to occur. Compare the following two examples:

128. *philm=te* absiph=oʃ. *Unmarked construction*
 film=OBL start=ACT.PST.1SG
 ‘I started the film.’

absiʔb ‘start’ in predicative function always appears in its unmarked use in the active. It can, however, also appear in the generic middle to denote that it took longer than usual to start the film, e.g., due technical problems, etc.

129. *philm=te* absiʔb=ki=j. *Generic middle*
 film=OBL start=MD.PST=1SG
 ‘I started the film (but it took me a long time).’

As *absiʔb* is bisyllabic, its masdar is identical to the underlying contentive morpheme. Thus, when *absiʔb* is the semantic base, the only indication that this is the marked construction is the use of the middle voice as opposed to the active.

- The generic middle can also denote a remote past. Consider the following examples with *hakɽe*, which in the unmarked construction is restricted to the active:

130. *goreʔj* hakɽe=yoʔ. *Unmarked construction*
 ox grunt=ACT.PST
 ‘The ox grunted.’ (unmarked)

131. *goreʔj* hakɽe=ki. *Generic middle*
 ox grunt=MD.PST
 ‘The ox grunted.’ (a long time back)

- Similarly, with respect to future events, the marked construction signals remote future time or future events whose time of occurrence is uncertain.

The generic middle can thus be used to express habituality, longer-than-normal duration, remote past, and remote or indefinite future. The unmarked use is underspecified with respect to these features, e.g., the unmarked construction is compatible with modifiers expressing 'over and over', 'long ago', etc., but this interpretation is not inherent in the unmarked usage.²¹

Although it was not possible to consult the following speaker from north Orissa on this construction, the following example from a monologue of his suggests that the Orissa dialect is much closer to the Simdega dialect in this respect. The semantic base of the TAM/PERSON-syntagma in the following example, *ikon*, when used in the unmarked construction, is restricted to the active, e.g. *ikon=o?=may* 'they did (once or more, just now or long ago, .)', whereas the semantic base in the following example of the generic middle is a masdar which is identical in form to the underlying contentive morpheme:

132. *soub apan apan ikon=ki=may.*
 all REFL REP do=MID.PST=3PL
 'All did [this] on their own.' [MS, 2:14]

This example is interesting as we have a TAM/PERSON-syntagma with a bisyllabic contentive morpheme in the generic middle, showing that this speaker's dialect, which is located just a few hours to the south of Simdega city, also allows polysyllabic contentive morphemes to appear in the generic middle. Furthermore, note the semantics of the predicate of the example: In his narrative, the speaker has just noted that the sons (= *all* in (132)) have collected their things, left, and now returned. Here, he is stressing the fact that each did this on his own. The use of the generic middle here thus differs somewhat from that discussed above, as

²¹ This category is semantically quite close to a category found in many Austronesian languages and referred to by Brill (2005: 32) as "actions without an endpoint". This group includes "tentative, aimless, or dispersive actions, or aspectually unbounded (non-completed, durative or iterative) actions, correlating with the notion of non-fully affected patients and consequently to a lower degree of transitivity and to intransitivization." As the function of this category seems to differ slightly in Kharia, including e.g., remote past / future but not including "non-fully affected patients", I will refer to it by the term generic, which is somewhat more neutral.

we are not dealing here with habituality or a remote past but rather with an event consisting of a multiplicity of individual actions. More research is necessary on this construction, as this function of the generic middle did not occur in conversations or interviews.

A number of contentive morphemes have taken on a slightly different meaning when used in the generic middle, although the origin of the new meaning from habituality / iterativity, etc., of the underived use is always apparent. A few examples are given in Table 6.22.

Table 6.22: Semi-lexicalized meanings in the generic middle

<i>Underlying contentive morpheme</i>	<i>Unmarked meaning (active)</i>	<i>Generic middle</i>
<i>bhaʔpu</i>	'boil (eggs)'	'overboil'
<i>lam</i>	'search for'	<i>lam-lam</i> 'hunt; search for a long time'
<i>paham, phaham, thaham</i>	'ponder, plan; decide'	'be uncertain as to what to do'

Finally, the generic middle is sometimes indicated by the form of the contentive morphemes themselves. This (very seldom) use is restricted in our data to contentive morphemes which end in *-ay* but which do not otherwise have the *-ay* / *-e* distinction (6.3.3) and which are restricted to the middle voice. In these cases, the (otherwise non-existent) form in *-e* can be used in the speech of some speakers to signal the marked construction.²²

133. *kuʔmuʔay* Active: –; Middle: 'get angry' *kuʔmuʔe* Middle: GENERIC

C. Middle: self-directed, Active: outwardly directed The semantics of this differential function, although closely related to both transitivity and reflexivity, must be considered separate from those two categories. Since some contentive morphemes in this category, such as *guʔuŋ* 'wash (feet)', have a direct object in both the middle and active, i.e., *kaʔa* 'feet', both members of these pairs may be transitive, while neither can be said to be reflexive in the true sense of the word, where agent and patient are

²² Note that this is not the same class of contentive morphemes as that discussed below in sections H and I.

identical. The difference in the case of, e.g., *gujuj* pertains to the possessor of the feet, i.e., one's own or someone else's feet.²³ Furthermore, as discussed in 6.5.1, Kharia has both a morphological and a syntactic reflexive construction.

Table 6.23 gives a few examples from our data. Note that most members of this class can be classified as "grooming" contentive morphemes, although not all.

Table 6.23: Self-directed vs. outwardly directed

	<i>Middle</i>	<i>Active</i>
<i>amo'd</i>	'wash one's own face' (TR)	'wash someone else's face' (TR)
<i>anargi</i>	'brush one's own teeth' (TR)	'brush someone else's teeth' (e.g. of a child) (TR)
<i>buli'j</i>	'scrub one's own head' (TR)	'scrub someone else's head' (TR)
<i>gujuj</i>	'wash one's own feet' (TR)	'wash someone else's feet' (TR)
<i>gunay</i>	'regret (for one's self)'	'regret (for another)'
<i>gu?the</i>	'wash one's own hands' (TR)	'wash someone else's hands'
<i>koy</i>	'shave (for some speakers only oneself, for others neutral)'	'shave someone else (only for some speakers)'
<i>kublāb</i>	'rinse out one's own mouth' (TR)	'rinse out s.o. else's mouth' (TR)
<i>khuray</i>	'shave oneself'	'shave someone else'
<i>paṅkha</i>	'fan oneself'	'fan someone else'
<i>uwa?</i>	'bathe' (ITR)	'bathe' (TR)

In a number of cases, the *-e / -ay* distinction in the morpheme itself combines with the active / middle distinction to express a similar semantic opposition:

134. <i>poṭhay</i>	Active: 'read to s.o.'	<i>poṭhe</i>	Middle: 'read (to oneself); study'
<i>pāūray</i>	Active: 'help s.o. swim'	<i>pāūre</i>	Middle: 'swim'

D. Middle: "Non-thorough", Active: "thorough" In this somewhat heterogeneous group, the notion is in some was less "thorough" when the

²³ It is customary in this region to wash a guest's feet when the guest is visiting either for the first time or has not been to visit for a while. This is either done by the eldest woman of the house or the eldest daughter.

TAM/PERSON-syntagma appears in the middle voice than when it appears in the active. Some examples are given in Table 6.24.

It is not clear at the moment to what extent these contentive morphemes form a coherent class. For example, while some, such as *alar*, *ju²q* and *thike*, refer to short-term events and situations (middle) as opposed to permanent situations or the end-result of an event (active), *cerberay* refers to a typical characteristic in the middle but to a specific action in the active and *camkay*, *kaŋi²j* and *lebui* appear to use the active to express “intensity” Further research is necessary here, especially since not all speakers were of the opinion that all the contentive morphemes referring to emotions can be used this way, although they did accept the other morphemes with the distinctions shown in the Table 6.24.

Table 6.24: “Non-thorough” vs. “thorough”

	<i>Middle</i>	<i>Active</i>
<i>alar</i>	‘love (for a short while)’	‘love (permanently)’
<i>dular</i>	‘like, feel love’	‘(show) love’
<i>camkay</i> ²⁴	‘warn’	‘scare’
<i>cerberay</i>	‘chirp’ (e.g. as a habit)	‘chirp’ (in reaction to something, e.g. as a warning)
<i>ju²q</i> ‘sprout roots’	‘sprout roots’	‘take root firmly / permanently’
<i>kaŋi²j</i> ‘believe’	‘believe, have faith in (more or less)’	‘believe in (strongly)’
<i>lebui</i>	‘love (somewhat)’	‘love (strongly)’
<i>thike</i>	‘last’ (neutral / not permanent)	‘last (permanently)’

E. Middle: Sudden / unexpected / non-volitional, Active: non-sudden / expected / volitional With contentive morphemes in this group, we generally find an intransitive (MIDDLE) / transitive (ACTIVE) distinction coupled with a sudden (MIDDLE) / non-sudden (ACTIVE) distinction, although some active forms are intransitive and some middle forms are transitive. This distinction can also be exploited to denote that the event referred to is depicted as unexpected in the middle voice, while the active denotes an expected, volitional action or is unmarked in this regard. Table 6.25 presents a few examples.

²⁴ This is only in reference to the base *camkay*, not the base *camke*, which means ‘become scared’

Table 6.25: Unexpected vs. expected

	<i>Middle</i>	<i>Active</i>
<i>gargaray</i>	'clear of the throat' (unintentional)	'gargle; clear the throat' (intentional)
<i>gunah</i>	'become guilty' (unknowingly)	'commit a sin'
<i>huřmuřay</i>	'bump into someone (by accident) (ITR)'	'bump into someone on purpose (ITR); block someone's way (TR)'
<i>kui</i>	'stumble upon'; 'appear', 'be found'	'(look for and) find'
<i>malum</i>	'come to know (accidentally)'	'find out (by searching)'
<i>onđor</i>	'hear (accidentally)'; 'be heard'	'(listen and) hear'
<i>yo</i>	'catch a glimpse of'; 'be seen'	'(look and) see'

The notion of unexpected vs. expected is, however, only tendential in this class; both the active and the middle are completely compatible with modifiers such as *koř-koř=ga* 'know-RDP=FOC' 'on purpose, knowingly' as well as *enem koř=ga* 'without know=FOC' 'by accident, without knowing', although there is a strong tendency to interpret the middle as non-volitional. Although speakers often indicated that the use of the middle with these contentive morphemes is "more sudden" than that of the active, we interpret this as meaning that this "suddenness" is primarily due to the unexpectedness of the event signalled by the middle voice. Further study is necessary here.

The following two examples, both taken from the same story, illustrate these subtle differences: The first depicts a sudden, unexpected event and the predicate, which appears in the middle voice, is intransitive. In the second example, we find a deliberate, non-sudden event being carried out, and the active-marked predicate is transitive.

Middle: sudden, unexpected, intransitive

In the following example, the soldiers have set out to search for the source of smoke in the forest which they have noticed while hunting. They are thus expecting to find a camp-fire. However, much to their surprise, they stumble upon a castle in the middle of the forest:

135. *jab* *ho=ki* *lam=na* *col=ki=may* *la?*
 CREL:TEMP that=PL seek=INF go=MID.PST=3PL then
- ho=ki=te* *kinir* *mo?jhi=te* *moř* *koleř=a?* *mohol*
 that=PL=OBL forest middle=OBL one king=GEN castle

kui=ki

find=MID.PST

‘When they went to look for [the source of the smoke], they [unexpectedly] found a king’s castle in the middle of the forest.’ [literally: a king’s castle was found to them] [HJPa:40]

Active: non-sudden, expected, transitive

In this example, a man asks for help in looking for his axe. Thus, the act of finding the axe in this example is the result of a deliberate, non-sudden action:

136. *la?* *ijn=te* *alsi* *lam=na* *buŋ* *saŋghar=e=pe* *oɔo?* *ber*
 then 1SG=OBL axe seek=INF INST help=ACT.IRR=2PL and who

kuy=e, *ho=kaŋ=te* *ijn=a?* *beŋi=ŋ=te*
 find=ACT.IRR that=SG.HUM=OBL 1SG=GEN daughter=1SG=OBL

ter[=e]=ijn.

give[=ACT.IRR]=1SG

‘Then [please] help me with looking for my axe, and *whoever finds* [it, I] will give him my daughter [in marriage].’ [HJPa:39]

F. Middle: no action, Active: action (SG); Middle: long duration, Active: shorter duration (DU, PL) There is only one morpheme in this class, the defective form *a* / *ani* ‘come on!’, which is restricted to the irrealis and the second-persons. The form *a* is found only in the singular, while *ani* is found in the dual and plural when unmarked for basic voice or when *ani* is the first sub-base in complex semantic bases where only the final semantic sub-base is marked for TAM/basic voice and person (6.3.2). The following discussion focuses on this defective morpheme when it is marked for person.²⁵

In the singular *a* is always marked for middle or active voice. The middle denotes that the addressee is already standing near to the speaker and need not “do” anything: *a=na!* in such cases is merely a kind of interjection, such as “OK!”, “Well then!”, etc., and simply denotes that something is about to happen, and the active form *a=e!* is used when the addressee must first come to the speaker before the next action occurs. This is summarized in Table 6.26.

²⁵ It should be noted that the use of this morpheme is somewhat speaker-specific. Further research is necessary.

Table 6.26: No action vs. action

<i>Middle</i>	<i>Active</i>
<i>a=na!</i> 'come on! (addressee is standing next to speaker)!'	<i>a=ye!</i> 'come on! (addressee must first come to the speaker)'

The dual and plural have three forms and function somewhat differently. Here, the base for middle and active voice is *a*, similar to the singular, but in the dual and plural the form may also be unmarked for basic voice, in which case it has the form *ani* and is followed only by the markers of the second person dual or plural.

In the middle voice, the addressees are typically still sitting and must first stand up and come to the speaker before the following action is to occur, whereas the active voice typically denotes that the addressees are already standing and must simply first come to the speaker. The form unmarked for basic voice, *ani*, then denotes that the addressees are already standing together with the speaker. Table 6.27 summarizes this.

Table 6.27: Delayed reaction (middle) vs. prompt reaction (active) vs. no action (unmarked)

<i>Middle</i>	<i>Active</i>	<i>Neither middle nor active</i>
<i>a=na=bar!</i> (DU) <i>a=na=pe!</i> (PL) 'come on (addressees are still sitting), let's go!'	<i>a=ye=bar!</i> (DU) <i>a=ye=pe!</i> (PL) 'come on (addressees are already standing), let's go!'	<i>ani=bar!</i> (DU) <i>ani=pe!</i> (PL) '(we're already standing together) let's go!'

Note that in the dual and plural, the "interjective" function of the middle voice of the singular, in which no action is indicated, is taken over by the form which is unmarked for basic voice and marked only for the addressee. This allows for the active / middle distinction to be exploited in a different way in the dual and plural than in the singular.

G. Middle: non-individuated object, Active: individuated object At present there is only one example in our data for this function, given in Table 6.28, although it is likely that more will turn up as research progresses.

Table 6.28: Non-individuated vs. individuated object

	<i>Middle</i>	<i>Active</i>
<i>peyar</i>	‘love (e.g. people in general)’	‘love (a particular person)’

The following two sections deal with speaker-specific uses in which contentive morphemes with two different forms (*-e* vs. *-ay*, cf. 6.3.3) are exploited in a way similar to the use of middle and active marking elsewhere. Thus, here it is not the portmanteau TAM/basic voice marking which is relevant, but rather the morphemes themselves. However, as these morphemes are otherwise restricted to the middle (in *-e*) or the active (in *-ay*), they may conveniently be dealt with here under basic voice, to which they are closely connected.

The following distinctions are generally found in environments where the use of the active or middle is required due to the presence of other morphological marking, e.g. the reciprocal, which requires the middle voice, or the causative, which requires the active (6.3.5).

H. *-e* Thwarted attempt, *-ay* successful action In this function, the use of the form of the morpheme ending in *-e*, which is otherwise restricted to the middle, denotes that an action could not be carried out, while the use of the form in *-ay*, which is otherwise restricted to the active, denotes that the actor was successful. Table 6.29 gives one example from our data. This function is highly speaker-specific and was not accepted by all speakers.²⁶

Table 6.29: Failure vs. success

<i>Form in the middle voice</i>	<i>Form in the active voice</i>
<i>bagre</i> ‘try to destroy’ (unmarked meaning: ‘become destroyed’)	<i>bagray</i> ‘destroy’
<i>kol bagre=ki=kiyar</i>	<i>kol bagray=ki=kiyar</i>
REC be.destroyed=MID.PST=DU	REC destroy=MID.PST=DU
‘they (DU) tried unsuccessfully to destroy something they own’	‘they (DU) destroyed something they own’

²⁶ The possessive interpretation in this example is expressed by the reciprocal marker (6.3.5).

This alternation is primarily found in environments in which a particular voice is required, as in Table 6.29, where the use of the reciprocal requires the middle voice. It is also occasionally found in other environments in combination with the active and middle: E.g. *thane* (MIDDLE) means 'try unsuccessfully to assert' while *thanay* (ACTIVE) means 'assert (successfully)'

I. *-e*: Indirect participation, *-ay*: direct participation In this function, which in our data is restricted to TAM/PERSON-syntagmas marked for the causative, the use of the form in *-e* indicates that the highest agent does not directly participate in the action but merely has it carried out by someone else, whereas the form in *-ay* indicates that the highest agent also participates in carrying out the action. Again, as in (H) above, it is not the portmanteau TAM/basic voice marking which signals this difference but rather the form of the contentive morpheme itself. An example from our data is presented in Table 6.30. This usage was not accepted by all speakers.

Table 6.30: Non-direct vs. direct participation

<i>Form in the middle voice</i>	<i>Form in the active voice</i>
<i>bagre</i> 'be(come) destroyed'	<i>bagray</i> 'destroy'
<i>ba<sup>?b>gre=yo?</i>	<i>ba<sup>?b>gray=o?</i>
be.destroyed-<CAUS>=ACT.PST	destroy-<CAUS>=ACT.PST
's/he had someone destroy'	's/he had someone destroy'
Subject does not participate in action	Subject also participates in action

Table 6.31 summarizes the respective differential functions of the middle and active. The criteria for (F)–(I) are in parentheses as they are either only valid for a small number of morphemes ((F)–(G)) or are highly speaker-specific ((H)–(I)).²⁷

²⁷ Note that this list closely resembles the list of factors given in Hopper & Thompson (1980) as contributing to "transitivity", although there are also interesting differences.

Table 6.31: Differential functions of Middle and Active

	<i>Middle</i>	<i>Active</i>
A	INTRANSITIVE	TRANSITIVE
B	GENERIC = habitual, long duration, remote past, remote / indefinite future	UNMARKED
C	SELF-DIRECTED	OUTWARDLY DIRECTED
D	“NON-THOROUGH”	“THOROUGH”
E	SUDDEN / UNEXPECTED / NONVOLITIONAL	NON-SUDDEN / EXPECTED / VOLITIONAL
F	(NO ACTION, S) / (LONG DURATION, DU/PL)	(ACTION, S) / (SHORTER DURATION, DU/PL)
G	(NON-INDIVIDUATED OBJECT)	(INDIVIDUATED OBJECT)
H	(THWARTED ATTEMPT)	(SUCCESSFUL ACTION)
I	(INDIRECT PARTICIPATION)	(DIRECT PARTICIPATION)

In the following section, the inherent “functions” of the middle and active are discussed, or rather the distribution of the middle and active voices with contentive morphemes which occur either with the middle or with the active, but not both (excluding the generic function of the middle, which holds for all contentive morphemes). Although properly speaking we are not dealing here with “functions” but only with tendencies, this label will be retained for convenience.

6.4.2.3 *Inherent functions*

The distribution of the middle and active with morphemes which may not appear with markers of both voices is based on the same principles as the differential functions of the active and middle. The inherent functions differ from the differential functions, however, in that the distribution of the inherent functions depends on the “prototypical” meaning of the contentive morpheme, whereas the differential functions refer to the particular meaning in a concrete instance.

- Middle-only contentive morphemes Table 6.32 presents a number of representative contentive morphemes which are restricted to the middle voice. These forms do not nicely fit into any one semantic or morphosyntactic category: While the large majority are intransitive, a number are inherently transitive and others may be used either transitively or intransitively. Also, although most denote avolitional events,

i.e., events for which volitionality plays no role, many are clearly non-volitional (at least generally), others are typically agentive, while others still may be either agentive or non-agentive. In the following discussion, these factors will be discussed individually.

Table 6.32: Middle-only contentive morphemes

<i>are</i> 'descend, go down'	<i>gur</i> 'fall'
<i>ajo'd</i> 'dry up (e.g. in a well)'	<i>hoy</i> 'become'
<i>aw</i> 'live, stay, QUAL'	<i>kamu</i> 'work' (ITR/TR)
<i>ba'j</i> 'like' (TR)	<i>kayom</i> 'speak' (ITR/TR)
<i>ba?khin</i> 'explain' (TR)	<i>la?</i> 'seem; EMOT'
<i>beloŋ</i> 'ripen'	<i>la'd</i> 'bake (chapatis)' (TR)
<i>botoŋ</i> 'become afraid, fear'	<i>leŋ</i> 'fly'
<i>del</i> 'come'	<i>sanŋo'd</i> 'go'
<i>ge'b</i> 'catch fire/burn'	<i>thartharay</i> 'tremble'
<i>go'j</i> 'die'	<i>urumda?</i> 'sweat'

Transitivity As Table 6.32 shows, almost all middle-only contentive morphemes are intransitive. Hence, we may consider intransitivity the primary inherent function of this class. In 6.4.2.2 above, it was argued that intransitivity is the primary differential function of the middle whereas the active is transitive. Intransitivity is thus both an inherent and a differential function of the middle voice.

Nevertheless, there are a few middle-only contentive morphemes which can take a direct object. Table 6.33 summarizes those from Table 6.32.

Table 6.33: Middle-only transitives

<i>ba'j</i> 'like' (TR)	<i>kayom</i> 'speak (a language, etc.)' (ITR/TR)
<i>ba?khin</i> 'explain'	<i>la'd</i> 'bake (bread, chapatis)'
<i>kamu</i> 'work' (ITR/TR)	

However, these are not "prototypical" transitives: E.g. *ba'j* 'like' actually describes more of a state than an action, whereas the remaining four contentive morphemes are quite restricted as to the identity of their possible objects. For example *kamu* 'work', when used transitively, can only take the cognate object *kamu* 'work':

137. (*kamu*) *kamu=ki=ŋ* 'I worked'
 work work=MID.PST=1SG

Similar restrictions hold for the remaining three morphemes: *laʔq* ‘bake (bread, chapatis)’ can only occur with *koloŋ* ‘bread, chapatis’ as its object. Similarly, if *kayom* ‘speak’ takes an object, this object is always a language name, e.g. *khaŋiya kayom* ‘speak Kharia’ *baʔkhin*, while less restricted, can nevertheless only take “abstract” objects, such as *kahani* ‘story’, *kayom* ‘matter’, etc. In other words, these five morphemes are low in “transitivity” (in Hopper & Thompson’s (1980) sense) and are far removed from “prototypical” actions which involve, among other things, an individuated, affected object.

Genericity With respect to the “inherent aspect” or *Aktionsart* of middle-only contentive morphemes, the data in Table 6.32 also show that, with the exception of contentive morphemes denoting motion, such as *qel* ‘come’ or *sangoʔq* ‘go’, most morphemes are what are often termed “eventives”²⁸ in that they do not really describe an action. Rather, for the most part they describe a process or change of state. This is often an inherently durative event, such as *aw* ‘stay, remain; QUAL’, *baʔj* ‘like’, *thartharay* ‘tremble’ or *urum qaʔ* ‘sweat’. Another large group of contentive morphemes, in fact much larger than the “inherent duratives”, denotes non-induced changes of state. To this group we may add “postural” contentive morphemes, or more precisely, “change of posture” contentive morphemes such as *qoko* ‘sit (down)’, as these can denote changes of state, although they are generally volitional. Table 6.34 presents a number of contentive morphemes from Table 6.32 and several others belonging to this class.

Table 6.34: “Eventive” middle-only contentive morphemes: Processes and changes-of-state

<i>ajoʔq</i> ‘dry up (e.g. in a well)’	<i>kamu</i> ‘work’ (ITR/TR)
<i>aw</i> ‘live, stay, QUAL’	<i>kayom</i> ‘speak’ (ITR/TR)
<i>baʔj</i> ‘like’ (TR)	<i>kosor</i> ‘dry up’ (ITR)
<i>beloŋ</i> ‘ripen’ (ITR)	<i>kulaŋ</i> ‘turn over’ (ITR)
<i>beroʔq</i> ‘rise’	<i>kulqaʔ</i> ‘get/have a fever’
<i>betoʔq</i> ‘become hungry’	<i>khisay</i> ‘become angry’
<i>betoʔq qaʔ</i> ‘become thirsty’	<i>laʔ</i> ‘seem; EMOT’
<i>bone</i> ‘become’	<i>laʔq</i> ‘bake (chapatis)’ (TR)

²⁸ This term as I use it corresponds (at least very closely) to those aspectual categories referred to as “noneventuality” by Klaiman (1991).

Table 6.34 (*cont.*)

<i>borol</i> 'live'	<i>leŋe</i> 'flow, float' (ITR)
<i>botoŋ</i> 'become afraid'	<i>lere?</i> 'become happy; rejoice'
<i>bul</i> 'get drunk, poisoned'	<i>leru</i> 'rest (only used for people)' (ITR)
<i>qoko</i> 'sit (down)'	<i>lo'b</i> 'get burned (of people), burn oneself'
<i>eŋhe'd</i> 'become bored'	<i>paŋom</i> 'regain strength'
<i>ge'b</i> 'catch fire/burn'	<i>si'd</i> 'become lost'
<i>gita?</i> 'lie (down)'	<i>siray</i> 'expire, die'
<i>go'j</i> 'die'	<i>su'd</i> 'become wet'
<i>gur</i> 'fall'	<i>tomon</i> 'stand (up)' (ITR)
<i>hoy</i> 'become'	<i>thartharay</i> 'tremble'
<i>isin</i> 'to cook (of rice)' (ITR)	<i>urumqa?</i> 'sweat'

Of course virtually all of the contentive morphemes shown in Tables 6.32 and 6.34 are intransitive, and it could be argued that valency alone is the decisive criterion. As we shall see below, however, it is not intransitivity alone which is responsible for this distribution as we also find intransitives which appear only with the active voice and which differ from these "eventives" in that, e.g., they do not refer to a long, drawn-out process or change of state but rather to a punctual action, such as *qera* 'take up lodgings' (ACTIVE), as opposed to *aw* 'live' (MIDDLE). "Eventive" is thus considered to be an inherent function of the middle voice. This corresponds closely to the generic differential function of this voice with respect to durativity.

Volitionality Ignoring here the five (potentially) transitive middle-only contentive morphemes in Table 6.32, the remaining contentive morphemes can be broken down into three groups on the basis of their typical volitionality status: Typically avolitional, typically volitional, and equally compatible with volitional or avolitional interpretations.

The contentive morphemes listed in Table 6.35, like almost all middle-only intransitives, are typically avolitional, i.e., volitionality is irrelevant with these morphemes. Again, although these contentive morphemes are all intransitive, there is more at issue here than mere intransitivity. Intransitives which are typically volitional, with the exception of most contentive morphemes of directed motion, are typically active-only, such as *dhãŋ* 'run, hurry', *yar* 'flee' In fact, we know of no active-only intransitive contentive morpheme which is not typically volitional. Hence avolitionality can be considered an inherent function of middle, as opposed to the potential volitionality of the active (see below).

Table 6.35: Middle-only, typically avolitional

<i>ajoʻd</i> ‘dry up’ (ITR)	<i>gur</i> ‘fall’
<i>aw</i> ‘live; stay; QUAL’	<i>hoy</i> ‘become’
<i>beloŋ</i> ‘ripen’	<i>kulda?</i> ‘have/get a fever’
<i>betoʻd</i> ‘become hungry’	<i>leŋe</i> ‘flow, float’
<i>betoʻd da?</i> ‘become thirsty’	<i>loʻb</i> ‘get burned (of people)’
<i>boŋ</i> ‘come to an end’	<i>puqu</i> ‘explode’
<i>borol</i> ‘live, be alive’	<i>siʻd</i> ‘become lost’
<i>bhore</i> ‘fill up (ITR)’	<i>siray</i> ‘expire, die’
<i>geʻb</i> ‘catch fire/burn’ (ITR)	<i>thartharay</i> ‘tremble’
<i>goʻj</i> ‘die’	<i>urum da?</i> ‘sweat’

There are also a number of middle-only contentive morphemes which are typically volitional and primarily denote directed motion and (change of) posture. They are thus an exception as they are middle-only yet potentially volitional. The fact that they are restricted to the middle voice must then be attributed to the primary inherent function of the middle, namely intransitivity. A few examples of these morphemes are given in Table 6.36.

Table 6.36: Middle-only, typically volitional

<i>buli</i> ‘wander’	<i>gita?</i> ‘lie (down)’
<i>qeʻb</i> ‘climb, mount’	<i>leru</i> ‘rest (of people)’
<i>qoko</i> ‘sit down’	<i>tomon</i> ‘stand (up)’

Finally, the contentive morphemes presented in Table 6.37 are equally compatible with both situations involving potential volitionality (= volitional and non-volitional) as well as avolitional situations with an inanimate subject incapable of volitionality. Again, the fact that these are restricted to the middle voice would appear to be based primarily on the fact that they are intransitive.

Table 6.37: Potentially volitional or avolitional middle-only contentive morphemes

<i>are</i> ‘descend’	<i>kulaŋ</i> ‘turn over’ (ITR)
<i>beroʻd</i> ‘rise’	<i>leʻd</i> ‘hide (ITR), become hidden’
<i>col</i> ‘go’	<i>leŋ</i> ‘fly’
<i>dam</i> ‘arrive’	<i>lere?</i> ‘rejoice; become happy’
<i>del</i> ‘come’	<i>mu?</i> ‘emerge’
<i>qiyar</i> ‘enter’	<i>sangoʻd</i> ‘go’
<i>eŋ</i> ‘return’ (ITR)	<i>sitiŋ</i> ‘move (ITR) to one side’

As the last two groups show, although (inherent) “eventiveness” and avolitionality can be considered inherent functions of the middle voice, these are secondary to intransitivity, which still must be considered the underlying or “unmarked” characteristic of this voice, although it can be overridden by other factors, as we shall also see in the following section.

In summary, middle-only contentive morphemes have the following three inherent functions:

- intransitivity
- eventiveness, i.e., inherently durative actions, changes of state
- tendentially avolitional

Note that this is actually a sub-set of the differential functions of the middle given in 6.4.2.2 above.²⁹ The differential functions of the middle which we do not find among the inherent functions are self-directed actions, thwarted attempt, non-direct involvement of the highest agent, and delayed or no action. We also do not find some functions of the generic middle, such as repeated actions or remote past, nor do we find non-volitionality in opposition to volitionality.

These functions are lacking here for the simple reason that they are only relevant within a productive system of oppositions, which is not possible when contentive morphemes can only occur in the middle voice. For example, failure to achieve an action presupposes an unmarked case in which an actor successfully performs the action denoted by the contentive morpheme. Thus failure to perform an action must be identifiable as a marked scenario, which is not possible in a system without oppositions. Similar comments hold for the other differential functions which are lacking here.

- Active-only contentive morphemes We now take a closer look at those contentive morphemes which only appear with active markers, ignoring here the (totally productive) temporo-aspectually marked function of the generic middle.

Transitivity

By far the largest group of active-only contentive morphemes consists of inherently transitive or ditransitive contentive morphemes. A representative

²⁹ Actually, the middle is tendentially non-volitional and the active tendentially volitional with the differential functions, whereas with the inherent functions, the middle is tendentially avolitional while the active is typically (potentially) volitional. Nevertheless, in both types of function, the middle denotes “reduced” volitionality.

list is given in Table 6.38. This is in line with the differential functions of the active and middle as well as with the fact that the primary inherent function of the middle is to denote intransitivity. Hence, transitivity is both an inherent and differential function of the active.

Table 6.38: Transitive active-only contentive morphemes

<i>biʔd</i> ‘pour out’	<i>oʔj</i> ‘take out’
<i>bui</i> ‘keep, raise (animals)’	<i>ol</i> ‘bring’
<i>coray</i> ‘steal’	<i>ragoʔj</i> ‘wash dishes’
<i>qoʔd</i> ‘take’	<i>soŋ</i> ‘buy’
<i>qul</i> ‘pour in’ (TR)	<i>su</i> ‘put on, wear’
<i>dhoʔ</i> ‘grab’	<i>ʔaŋgay</i> ‘hang up’
<i>gil</i> ‘beat (a person, pulse, etc.)’	<i>ter</i> ‘give’
<i>goʔ</i> ‘carry’	<i>til</i> ‘bury’
<i>jaʔb</i> ‘grab’	<i>tol</i> ‘bind’
<i>jiʔb</i> ‘touch’ (TR)	<i>tur</i> ‘shake out’
<i>juŋ</i> ‘ask’ (TR)	<i>uʔchi</i> ‘despise’
<i>kaquʔ</i> ‘hug’ (TR)	<i>uʔd</i> ‘drink’
<i>lam</i> ‘seek’	<i>un</i> ‘set’ (TR)
<i>noʔ</i> ‘eat’ (TR)	<i>uʔay</i> ‘waste’ (TR)

There is also a considerable number of intransitive active-only contentive morphemes, a sample of which is given in Table 6.39. These morphemes will now be dealt with with respect to volitionality and genericity.

Table 6.39: Intransitive active-only contentive morphemes

<i>aʔde</i> ‘stand firm, wait’	<i>iʔchōy</i> ‘fart’
<i>akil</i> ‘think (ITR/TR)’	<i>ij</i> ‘defecate’
<i>aŋaʔbqaʔ</i> ‘yawn’	<i>kaluwa</i> ‘eat dinner’
<i>bajaʔ</i> ‘vomit’	<i>kan</i> ‘fast’
<i>biʔthuŋ</i> ‘spit’	<i>khuʔ</i> ‘cough’
<i>chikay</i> ‘sneeze’	<i>laqa</i> ‘laugh’
<i>qera</i> ‘take up lodgings’	<i>lur</i> ‘reason (v.)’
<i>qhekar</i> ‘burp’	<i>nasta</i> ‘eat breakfast’
<i>dhāy</i> ‘run’	<i>paday</i> ‘fart’
<i>hāʔchiŋ</i> ‘sneeze’	<i>phikir</i> ‘think, reflect’
<i>haqa</i> ‘urinate’ ³⁰	<i>sēreʔ sēreʔ</i> ‘sniffle (when one has a cold)’
<i>hagay</i> ‘defecate’	<i>socay</i> ‘think (ITR/TR)’
<i>hakʔe</i> ‘grunt (of oxen)’	<i>yar</i> ‘flee’

³⁰ *haqa* ‘urinate’ can also be used transitively, but only with the cognate object *haqa* ‘urine’, a usage which however seems to be very seldom.

Volitionality

As Table 6.39 shows, many intransitive morphemes denoting body functions, such as *khu?* 'cough' and *chikay* 'sneeze',³¹ are active-only. Also, a number of morphemes denoting cognitive processes which would appear to be typically volitional, such as *akil* 'think', *phikir* 'worry, think about' *lur* 'reason', are active-only morphemes.

Let us now take a closer look at this group with respect to volitionality and (non-) eventiveness. For the sake of presentation, we begin with a discussion of active-only contentive morphemes and compare them with middle-only morphemes discussed above with respect to these two functions.

Consider first the volitional contentive morphemes in Table 6.40, all of which are active-only:

Table 6.40: Volitional, active-only contentive morphemes

<i>qera</i> 'take up lodgings'	<i>nasta</i> 'eat breakfast'
<i>dhãy</i> 'run, hurry'	<i>kan</i> 'fast'
<i>kaluwa</i> 'eat dinner'	<i>yar</i> 'flee'

Consider now the morphemes denoting body functions in Table 6.41, also active-only:

Table 6.41: Active-only, typically volitional body functions

<i>bi?thuŋ</i> 'spit'	<i>iʔ</i> 'defecate'
<i>haqa</i> 'urinate'	<i>laqa</i> 'laugh'
<i>hagay</i> 'defecate'	

That these are in fact treated as potentially volitional in Kharia is shown by the following example, in which the use of modifiers such as *koŋ-koŋ=ga* [KNOW-RDP=FOC] 'knowingly, deliberately' and *enem koŋ=ga* [without know=FOC] 'accidentally' is compatible with these contentive morphemes.

³¹ Speakers indicated in interviews that *chikay* 'sneeze' may also appear in the middle voice, with no apparent semantic difference, although in all instances I have come across, it is found in the active. For this reason, I treat it here as an active-only contentive morpheme, although this is somewhat tentative.

138. *ho=kaɾ koŋ-kon=ga / enem koŋ=ga haɖa=yo? /*
 that=SG.HUM KNOW-RDP=FOC without KNOW=FOC urinate=ACT.PST
biʔthuŋ=oʔ.
 spit=ACT.PST
 ‘S/he deliberately / accidentally peed / spit.’

The active-only morphemes in Table 6.42 would seem to be of a less volitional nature than those given in the previous two lists. However, while they are not as easily restrained as the actions of these last two groups, they can often be successfully suppressed for a certain amount of time. Also, the experiencer subject is generally aware that the action is about to take place and can decide whether the action is to be freely allowed or whether s/he should at least try to suppress it.

Table 6.42: Active-only, potentially volitional body functions

<i>aŋaʔbɖaʔ</i> ‘yawn’	<i>iʔchöy</i> ‘fart’
<i>bajaʔ</i> ‘vomit’	<i>khuʔ</i> ‘cough’
<i>chikay</i> ‘sneeze’	<i>paday</i> ‘fart’
<i>hãʔchiŋ</i> ‘sneeze’	<i>sëreʔ sëreʔ</i> ‘sniffle (when one has a cold)’

Again, this is supported by the following example, in which expressions denoting (non-) volitionality are compatible with these contentive morphemes.

139. *ho=kaɾ koŋ-kon=ga / enem koŋ=ga aŋaʔbɖaʔ=yo? /*
 that=SG.HUM KNOW-RDP=FOC without KNOW=FOC vomit=ACT.PST
khukh=oʔ / paday=oʔ.
 cough=ACT.PST fart=ACT.PST
 ‘S/he deliberately / accidentally vomited / coughed / farted.’

Consider now the morphemes in Table 6.43, all middle-only and many of which refer to body functions, all of which appear to be avolitional events.

Table 6.43: Avolitional middle-only contentive morphemes

<i>ajoʔd</i> 'dry up'	<i>goʔj</i> 'die'
<i>beloŋ</i> 'ripen'	<i>gur</i> 'fall'
<i>betoʔd</i> 'become hungry'	<i>kulɔa?</i> 'have / get a fever'
<i>betoʔd ɔa?</i> 'become thirsty'	<i>loʔb</i> 'get burned (of people), burn oneself'
<i>boŋ</i> 'come to an end'	<i>puɔu</i> 'explode'
<i>borol</i> 'live, be alive'	<i>siʔd</i> 'become / get lost'
<i>botoŋ</i> 'become afraid, fear'	<i>siray</i> 'expire, die'
<i>geʔb</i> 'burn/catch fire'	<i>urum ɔa?</i> 'sweat'

While the avolitional status of contentive morphemes which (at least generally) require an inanimate subject, such as *ajoʔd* 'dry up' or *beloŋ* 'ripen', is self-evident, this also holds for morphemes in this list such as *kulɔa?* 'have / get a fever' and *urum ɔa?* 'sweat', which require an animate subject. To put it another way, while we can often refrain from the actions in Tables 6.40–6.42, or at least try to suppress them, one cannot refrain from getting a fever or sweating—events such as these simply happen. Thus, one cannot even try to refrain from the body functions denoted by the middle-only morphemes, a fact which clearly distinguishes them from active-only morphemes denoting body functions. That these morphemes are in fact treated as avolitional in Kharía is confirmed by the following data:

140. *ho=kaŋ* (**koŋ-kon=ga* / **enem kon=ga*) *kulɔa?=ki* /
 that=SG.HUM KNOW-RDP=FOC without KNOW=FOC fever=MID.PST
urum ɔa?=ki.
 sweat=MID.PST
 'S/he (*deliberately / *accidentally) got a fever / sweat.'

These data justify the conclusion that potential volitionality is an inherent function of active-only contentive morphemes while avolitionality is inherent to middle-only contentive morphemes, excepting the morphemes of motion and change-of-posture.³²

³² Although the differential functions of the active and middle are largely the same as those discussed in Klaiman (1991), the fact that active-only morphemes are at least usually potentially volitional and middle-only morphemes largely avolitional would seem to argue quite strongly against applying Klaiman's analysis of middle-only predicates as

Non-“eventiveness” of the active

The tables above also indicate that active-only morphemes, regardless of their volitional status, have a strong tendency to be either activities, punctual actions or at least to be of much shorter duration than middle-only morphemes which, as we have seen, show a strong tendency to be duratives or spontaneous changes-of-state. This is especially true of the minimally controllable *baja?* ‘vomit’, *hã?chin?* ‘sneeze’ or *khu?* ‘cough’, as compared with inherently durative middle-only body functions such as *kulqa?* ‘have / get a fever’, *thartharay* ‘tremble’ and *urum qa?* ‘sweat’. This fits in well with the observations made above that eventiveness (or “genericity”) is both a differential and inherent function of the middle voice whereas the active is unmarked in this respect.

Summary of the inherent functions We can summarize the inherent functions of the middle and active voice as in Table 6.44.

Table 6.44: Inherent functions of the Middle and Active

<i>Middle</i>	<i>Active</i>
Intransitive	Transitive
Inherently long duration / Change of state	Unmarked
Avolitional	Potentially volitional

The “functions” given in Table 6.44 are clearly a sub-set of the differential functions of the middle and active voices and, as argued above, those differential functions which are not found among the inherent functions can be accounted for by the fact that it is in the nature of the inherent functions that no oppositions can be expressed.

While these criteria may not always suffice to predict the class to which a non-alternating contentive morpheme belongs, they do account for the attested data quite well and, in cases where there are semantically similar morphemes belonging to both classes, these criteria can often explain why a particular morpheme is grouped the way it is. Table 6.45 presents three such examples.

being “deponents”, in which the subject is the controller of the action, to Kharia. “Control” in Kharia is a typical feature of the active—not the middle—both in differential and inherent functions.

Table 6.45: Semantically similar middle-only and active-only morphemes

<i>Middle</i>	<i>Active</i>
<i>aw</i> 'remain, stay, live; QUAL'	<i>ɖera</i> 'take up lodgings, live (somewhere for a certain time)'
<i>borol</i> 'live; be alive'	
LONG DURATION, TYPICALLY	RELATIVELY SHORT DURATION, TYPICALLY
AVOLITIONAL	VOLITIONAL
<i>kayom</i> 'speak'	<i>gam</i> 'say'
1. STATE (I.E., ABILITY TO SPEAK A LANGUAGE): AVOLITIONAL, TRANSITIVE	ACTION: RELATIVELY SHORT DURATION, VOLITIONAL, TRANSITIVE
2. ACTION: VOLITIONAL, INTRANSITIVE	
BOTH MEANINGS: LONG DURATION	<i>yar</i> 'flee' (ANIMATE SUBJECT), 'take off quickly' (INANIMATE / ANIMATE SUBJECT)
<i>col</i> 'move, go', <i>buli</i> 'wander'	GENERALLY VOLITIONAL, SHORT DURATION
LONG DURATION, INTRANSITIVE	

Many questions remain with respect to the distribution of the middle and active voices with respect to their inherent "functions", and it may well turn out that exceptions to the general tendencies given here will eventually be found, although all of the active-only and middle-only contentive morphemes in our data seem to fit in well with the criteria given in the sections above.

Nevertheless, there remain a number of contentive morphemes which can appear in both the middle and active voices with no apparent semantic difference. Table 6.46 provides a few examples.

It remains to be seen whether semantic differences can be found for each of these contentive morphemes (as well as others). The use of active vs. middle with a large number of contentive morphemes is highly speaker-specific or typical of certain regions, and speakers' intuitions, at least in interviews, often differ considerably with respect to the "correct" usage and meaning of a particular morpheme in predicative function. In such cases, when differences of opinion have arisen among the speakers themselves, preference has been given to the data obtained from speakers who come from the general vicinity of the city of Simdega in the district of the same name, as this appears to be the region which shows the most productive use of these two categories. Further research is necessary.

Table 6.46: Active/Middle contentive morphemes with no apparent semantic difference

<i>akam</i>	'be absent'
<i>also?</i>	'pull up rice plants at harvest-time'
<i>aj</i>	'dawn; clear (of the sky)'
<i>dɔŋ</i>	'become sufficient'
<i>dure</i>	'coo (of pigeons)' (middle voice preferred)
<i>gonɔŋ</i>	'lament'
<i>hemalu</i>	'do slowly'
<i>isar</i>	'hate'
<i>jete^ʔb</i>	'drip; drizzle'
<i>kabrɔy</i>	'do', used only in a derogatory sense with respect to someone else's actions
<i>le^ʔb</i>	'smear, plaster'
<i>phusuʔuru</i>	'whisper'
<i>siskari</i>	'whistle'
<i>sonhor</i>	'eat' (of guests at a wedding feast after the wedding ceremony)

6.4.2.4 *A semantic map of the middle voice in Kharia*

Kemmer (1993) makes use of semantic maps to demonstrate the contiguity of the semantic areas covered by the middle voice in her cross-linguistic study. The structure of this mapping, an example of which is given in Kemmer (1993: 202, repeated here as Figure 6.1), is arranged as follows: At the top and bottom of this map we find (prototypical) two- and one-participant constructions, respectively, with both of these extremes typically appearing in the active voice. To the far right of the diagram, in the vertical middle, is the passive and opposite it, to the far left, are natural reciprocal events. In the middle of the map are direct reflexives, as Kemmer assumes these to occupy an intermediate position between prototypical one- and two-participant events (e.g., Kemmer, 1993: 73). The remaining central region of this map is then filled with various semantic areas often covered by the middle voice, arranged so that those areas which have most in common semantically are closer to one another.

Examples for some of these semantic areas, taken from Appendix A in Kemmer (1993: 267–270), are given with their Kharia equivalents in Table 6.47. This list is far from exhaustive and is only intended to give the reader unfamiliar with Kemmer's study an idea of what is meant by some of the terms. The data in Table 6.47 should also not be taken as meaning that all contentive morphemes in these semantic areas are obligatorily marked for the middle voice. For example, although almost

all intransitive contentive morphemes denoting “Translational Motion” (e.g., *col* ‘go’) appear exclusively in the middle voice, a small number nevertheless take the markers of the active, such as *yar* ‘flee’

By way of comparison, Figure 6.2 shows the approximate semantic range of the middle voice in Kharia, denoted by the shaded area. As direct reflexives, indirect reflexives and the passive (partially including Kemmer’s “Passive Middle”) are signalled by the same marker in Kharia, *dom* (6.5.1), these have been joined together in Figure 6.2. The areas “Reciprocal” and reflexive / passive have been set off in this map as they are only indirectly related to the middle voice in Kharia: These categories are expressed by separate marking: *kol* for the reciprocal and *dom* for the reflexive and passive, so that the middle voice is not the primary marker of these functions. However, the middle is connected to these semantic concepts in Kharia in that any TAM/PERSON-syntagma marked for one of these categories obligatorily appears in the middle voice.

The fact that the large majority of Kemmer’s “Cognition Middle” predicates appear in the active in Kharia (*koŋ* ‘know’, *socay* ‘think’, *iriʔb* ‘forget’, *konoʔd*, *paham* ‘think; remember’) is indicated in Figure 6.2 as a gap in the otherwise largely contiguous middle domain, which are termed here “Cognition Active” In addition, as a number of “natural reciprocal events” are always marked for the active, this semantic area is only half shaded in Figure 6.2.

Note that Kemmer differentiates between those intransitive categories listed in the lower half of Figure 6.1, such as “Change in Body Posture”, “Translational Motion”, etc., and (prototypical) “One-Participant Events” such as morphemes denoting ‘go’ (Kemmer, 1993: 204), because at least in most languages with a middle voice, most intransitive morphemes in fact require the active voice, whereas many of the other, more central categories in this figure are marked for the middle voice. To emphasize the fact that one-participant events in Kharia are generally marked for the middle voice, the term ACTIVE from the bottom of Kemmer’s map has been replaced by MIDDLE and this category has been incorporated into the shaded area in Figure 6.2. However, as there are a few, typically agentive, active-only intransitive contentive morphemes denoting translational motion, such as *yar* ‘flee’ and *dhãy* ‘run; hurry’, these have been added as a separate category, “Translational Active”, to emphasize the fact that they are exceptions to the general rule and do not represent the unmarked “One-Participant Events” Kemmer is referring to.

Table 6.47: Representative examples of contentive morphemes in the middle voice according to Kemmer's (1993) classification

Naturally reciprocal events and naturally collective actions	
<i>are</i> 'prepare to fight'	<i>kayom</i> 'talk, converse'
<i>katiʔb</i> 'come together, meet, assemble'	<i>metiʔb</i> 'come together, meet, assemble'
Grooming	
<i>amoʔd</i> 'wash one's own face' (TR)	<i>buliʔj</i> 'scrub one's own head' (TR)
<i>gujuj</i> 'wash one's own feet' (TR)	<i>uwaʔ</i> 'bathe' (ITR)
Nontranslational motion	
<i>kulaŋ</i> 'turn over' (ITR)	<i>hinduŋ</i> 'stoop, bend' (ITR)
<i>juŋ</i> 'stretch out (of legs)' (ITR)	<i>japaʔ</i> 'lean against' (ITR)
Change in body posture	
<i>ɖoko</i> 'sit (down)'	<i>tomon</i> 'stand (up)'
<i>gitaʔ</i> 'lie (down)'	
Translational motion	
<i>col</i> 'go'	<i>leŋ</i> 'fly'
<i>kabuŋ</i> 'crawl'	<i>buli</i> 'wander, stroll'
Emotion middle	
1-participant	
<i>lereʔ</i> 'become glad, rejoice'	<i>botoŋ</i> 'become frightened, fear'
2-participant	
<i>baʔj</i> 'agree, like, love'	<i>ghinay</i> 'hate'
Spontaneous events	
Motion	
<i>gur</i> 'fall'	<i>ɖeʔb</i> 'rise'
Associated with animate beings	
<i>goʔj</i> 'die'	<i>janam</i> 'be born'
Associated with inanimate beings	
<i>samay</i> 'sink'	<i>are</i> 'ascend'
<i>kuj</i> 'swell'	<i>konon</i> 'become small, shorten' (ITR)
<i>ɖeŋ</i> 'cook (of pulse, vegetables)' (ITR)	<i>gaʔ</i> 'rip, tear (of paper)' (ITR)
<i>puɖu</i> 'explode'	<i>rosoʔb</i> 'disintegrate'
<i>seʔ</i> 'fill (ITR); be contained, fit'	<i>ruʔ</i> 'open' (ITR)
<i>hoy</i> 'become'	<i>tuʔ, tuʔd</i> 'finish (ITR), come to an end'
<i>camke</i> 'flash, shine'	<i>laʔ</i> 'seem'

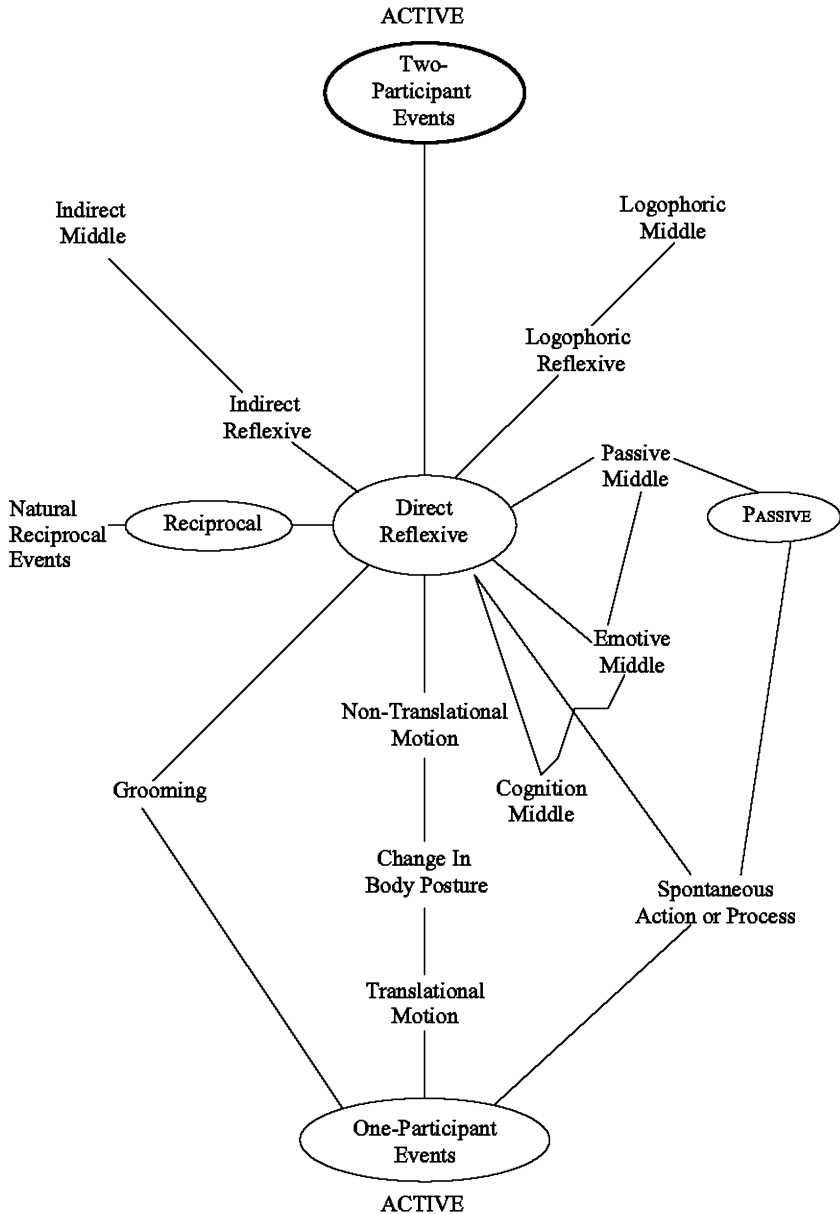


Figure 6.1: Semantic relations among middle and other situation types. From: Suzanne Kenner. *The middle voice*. 1993: 202. With kind permission by John Benjamins Publishing Company, Amsterdam/Philadelphia www.benjamins.com

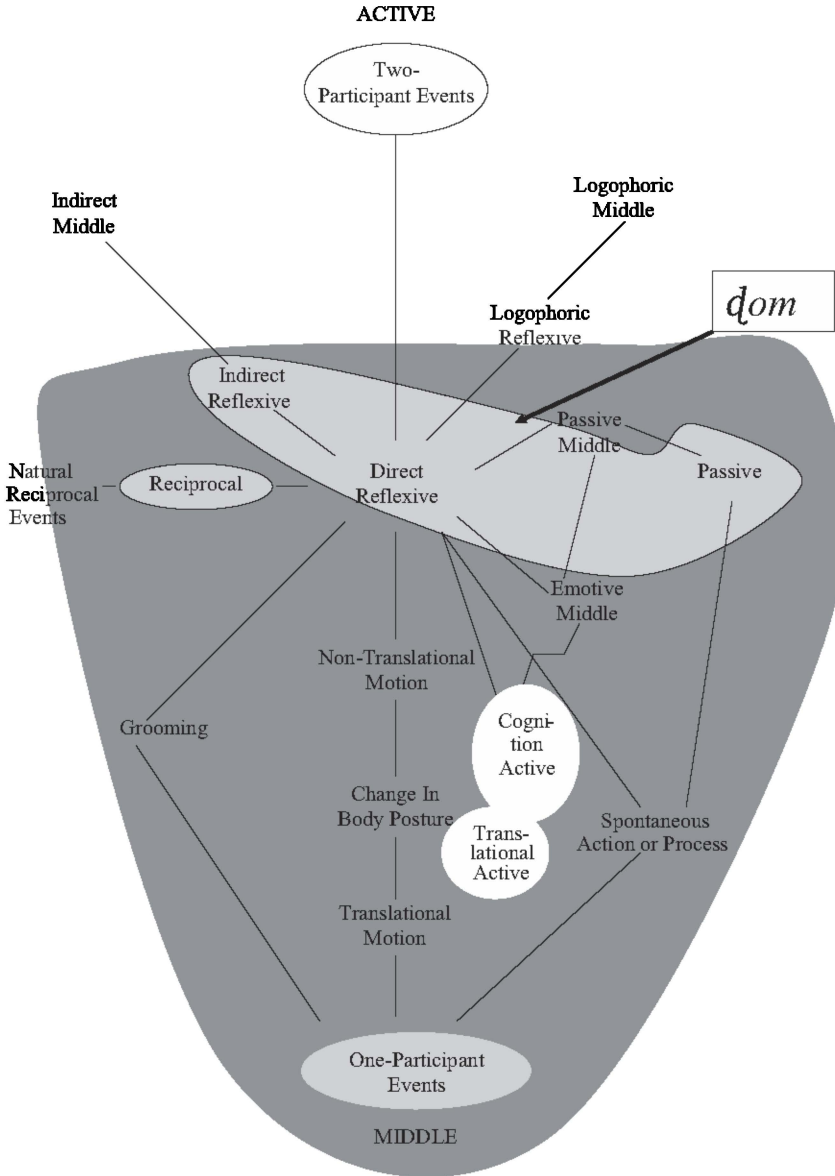


Figure 6.2: The semantic range of the middle in Kharia. Based on Suzanne Kemmer. *The middle voice*. 1993: 202. With kind permission by John Benjamins Publishing Company, Amsterdam/Philadelphia. www.benjamins.com

6.5 "v2s"—MARKERS OF THE PASSIVE / REFLEXIVE AND *AKTIONSPORT*

In this section those categories are dealt with which are referred to in Table 6.1 as "v2s" and which appear between the semantic base of the TAM/PERSON-syntagma and the perfect marker. The term "v2" has become more or less standard in South Asian linguistics although the term "explicator verbs" is also found. These units mark categories such as the passive/reflexive and *Aktionsart* in its broadest sense, including telicity, durativity and many other categories.

All of these morphemes are grammatical words, whereas their status as phonological words is somewhat unclear with respect to intonation (2.5). They may be considered phonological words, however, to the extent that they may be separated from the semantic base of the TAM/PERSON-syntagma as well as from each other by the floating pragmatic markers, such as =ga 'FOC' Cf. the following example:

141. *karay* go¹d=te=ga / *karay*=ga go²d=te
do C:TEL=ACT.PRS=FOC do=FOC C:TEL=ACT.PRS
's/he does the job'

A number of these "v2s" are also homophonous with contentive morphemes, and it seems reasonable to assume that they have all evolved from contentive morphemes.³³

6.5.1 *Passive / Reflexive: d̥om*

The passive/reflexive is marked by *d̥om*, which has no independent lexical meaning, although it is homophonous with the marker of 3rd person, inalienable possession (5.5). This category has two primary functions:

- It denotes a backgrounding passive, in which the agent of the action is suppressed or demoted to adjunct status.
- It denotes reflexivity, i.e., the subject performs an action on itself.

³³ The status of these "v2s" as derivational markers cannot be answered at this point. Many are highly restricted in terms of the morphemes with which they may combine. This, combined with the occasionally unpredictable resultant semantics (e.g., 6.5.9) suggests that they are derivational markers, although the fact that they are grammatical (and perhaps also phonological) words argues against this. As their status with respect to derivation is not of importance for a discussion of their use, this issue will not be dealt with further here.

Thus, the following TAM/PERSON-syntagma can have two very different meanings, depending on context:

142. *yo* *ḍom=ki=kiyar.*
 see PASS/REFL=MID.PST=DU
 ‘They (DU) were seen (by someone else).’ or
 ‘They (DU) saw themselves (e.g., in the mirror).’

In the passive use of this morpheme, the undergoer may appear either in the direct case, in which case it is “subject”³⁴ and is cross-referenced on the TAM/PERSON-syntagma (143), or it may appear in the oblique case, if definite, in which case the TAM/PERSON-syntagma has default marking, formally unmarked (= 3rd person (singular)) (144).

143. *pothi=ki ter ḍom=ki=may.* ‘(The) books were given.’
 book=PL give PASS=MID.PST=PL

144. *pothi=ki=te ter ḍom=ki.* ‘The books were given.’
 book=PL=OBL give PASS=MID.PST

When questioned as to the meanings of the two examples, speakers indicated that the first example may refer to books in general, whereas the presence of =*te* in the second example refers to specific books. In other examples, speakers indicated that the presence of the oblique marker on the patient meant that this was “more affected” than the construction in which the patient was the subject of the clause.

This category can also express an indirect reflexive meaning. In cases such as these, subject marking on the TAM/PERSON-syntagma can be an indication of which interpretation is intended, i.e., passive or reflexive. For example, in (145) the third-person, plural marking on the TAM/PERSON-syntagma is a clear indication that the passive interpretation is not intended, as the patient/object here, *komaŋ=te* ‘meat’ (OBJECT), is marked for the oblique case and cannot be interpreted as plural.

145. ... *taʔj ḍom=ki=may komaŋ=te.*
 distribute REFL=MID.PST=3PL meat=OBL
 they distributed the meat amongst themselves.’ [AK, 1:56]

³⁴ On the status of the “overt subject” as the subject of the clause, see Section 7.1.

In the absence of such explicit indicators, only context can decide which interpretation is intended.

Although all speakers we spoke with agreed that the reflexive interpretation of this morpheme is grammatical, examples such as (142) with a direct reflexive interpretation do not occur in our own corpus. However, two examples of this usage were found in a Kharia drama, both from the same page (Kerkeṭṭā, 1990: 12):

146. *koḍḍuru?*=ki *gam*=na=? *hosiyar* *gam* *ḍom*=ta=ki.
 man=PL say=INF=GEN intelligent say REFL=MID.PRS=PL
 'Men *call themselves* intelligent in (= of) speech.'

147. *apan* *dular* *konselḍug*=a? *mugamte* *koḍḍuru?*=ki *juḡhay*
 own love woman=GEN before man=PL much
ḍiṛḡgar ob-yo *ḍom*=ta=ki.
 brave CAUS-see REFL=MID.PRS=PL
 'In front of their beloved wives, men *show themselves* [to be] very brave.'

More commonly, the reflexive proform *apan*, which has been borrowed from Sadri, is used to denote reflexivity ((148), see Section 7.1.2 for further examples) or an object is used which is identical to the subject except that it is marked for the oblique case (149).

148. *aina*=te *apan*=te *yo*=yoʔj.
 mirror=OBL REFL=OBL see=ACT.PST.1SG
 'I saw myself in the mirror.'

149. *ijn* *ijn*=te *gil*=t[e]=ijn.
 1SG 1SG=OBL beat=ACT.PRS=1SG
 'I am beating myself.'

[Pinnow, 1966: 113, from Banerjee, 1894: 10]

The passive interpretation is also compatible with intransitive predicates. Here, the fact that the event was carried out BY SOMEONE is backgrounded and only the event itself is foregrounded, i.e., the fact that it took place.

150. *aw* *ɖom=ki*
 stay PASS=MID.PST
 ‘Someone stayed [but I don’t know who or it’s not important who].’

This same morpheme can also be used in response to polar or constituent questions. With polar questions, the answer to such questions often consists of a repetition of the TAM/PERSON-syntagma marked for the passive, although the TAM/PERSON-syntagma in the question itself did not appear in the passive. This use has long been noted for Kharia. E.g., Kullū (1981: 71f.) notes this use in answers in general and gives the following example (slightly adapted from Kullū, 1981: 72):

151. Q: *pe?* *ɲog=e=m?* ‘Will you eat rice (i.e. dinner)?’
 rice eat=ACT.IRR=2SG
- A: *ɲo?* *ɖom=na* ‘I will.’ (= ‘[It] will be eaten.’)
 eat PASS=MID.IRR

The following example for a constituent or “*wh-*” question is from interviews:

152. Q: *dilli tay lanɖan i=ghay col=ki=m?*
 Delhi ABL London what=way go=MID.PST=2SG
 ‘How did you go from Delhi to London?’
- A: *dilli tay lanɖan uɾanbaj buŋ leŋ ɖom=ki.*
 Delhi ABL London airplane INST fly PASS=MID.PST
 ‘By plane.’ (= ‘[It] was flown by plane from Delhi to London.’)

The following is an example for a follow-up question in the passive which is similar to the examples above. The morpheme *ikon* in the dialect of the speaker of the following example means ‘to do something or other somehow or other’³⁵ and signals that the speaker is still trying to

³⁵ For most speakers, *ikon* is a pause word meaning approximately ‘umh...’, which derives from the converbial form of *i* ‘(do) what?’, while for others, primarily older speakers, *ikon* simply means ‘do’ (from which the converbial marker *kon* would seem to have arisen (6.6.2.3)). The meaning ‘do somehow or other’ in this speaker’s dialect apparently

remember what he did, hence the follow-up question in the example. The contentive morpheme *karay*, borrowed from Indo-Aryan, does not have this connotation and simply means 'make, do'

153. A: *u kamu ikon=oʔj*
 this work do=ACT.PST.1SG
 'I did this work. (hold on, I'm still thinking of what it was).'

B: *i karay qom=ki?*
 what do PASS=MID.PST
 'What did you do?' (= 'What was done?')

The use of the passive in such constructions is clearly motivated by the fact that the subject (= S/A) of the expected response is generally known, unless it is this entity which is being questioned. As such, this entity can be backgrounded by means of the passive, as its identity is obvious from context. This use of the passive is not obligatory, however, and the active form of the TAM/PERSON-syntagma may also be used in responses.

No other Munda language except the closely related Juang seems to have a passive category (other than the intransitive function of the middle voice). The passive marker in Juang, *jim*, is cognate with the Kharia autopoesis marker *jom*, with both markers deriving from lexical morphemes meaning 'eat'³⁶ (Pinnow, 1966: 112). It is not, however, cognate with the Kharia passive / reflexive marker.

results from the original meaning 'do' but has been influenced by the meaning of the homophonous pause word.

³⁶ Pinnow (1966: 112) notes that these morphemes are both homophonous in their respective languages with contentive morphemes meaning 'eat', although all speakers I consulted rejected this for Kharia. Nevertheless, whatever the status of *jom* may have been at the time Pinnow gathered his data, comparative data from a number of other Munda languages clearly show that the autopoesis marker *jom* in Kharia does derive from a contentive morpheme originally meaning 'eat'

Note also that Pinnow (1966: 112) considers *jom* to be an allomorph of *qom*, a view which I do not share (see "Autopoesis", 6.5.10). He also considers *qom* to be only a passive marker, not a reflexive marker, and writes explicitly (Pinnow, 1966: 113) "In Kharia a special form for the reflexive no longer exists: . . .". However, he does note on the same page that a reflexive interpretation is possible with the "passive" marker *qom*, citing an example from Banerjee (1894: 10): *jn=ga gil qom=i[a]=jn* [1SG=FOC beat PASS=MID.PRS=1SG] 'I am beating myself' (transcription altered, my gloss). In explaining this form, he writes "Literally 'myself am being beaten'" where "myself" apparently refers to *jn=ga*. This suggests

As Pinnow notes, there is a cognate form in some South Munda languages but which in those languages is a reflexive, not a passive, category; cf. Juang *ɬom* / *rom*, Sora *dəm*. This suggests that the reflexive interpretation of *ɬom* was the original function of this category, as Pinnow (1966: 113) suggests (see also Anderson, 2007: 41). This conclusion is supported by the following facts:

- Despite the fact that the reflexive use of *ɬom* is so seldom in actual texts, all speakers we consulted were familiar with the (direct) reflexive function of this category and accepted it as grammatical;
- From a cross-linguistic perspective it is quite common for reflexive constructions to develop further into a passive.

It seems that, with the wholesale importation of the Indo-Aryan reflexive construction with *apan* mentioned above and dealt with in Section 7.1.2 in some detail, the reflexive use of *ɬom* has receded from this domain and is perhaps in the process of being lost, while the passive use is still common, which would also account for the cognate reflexive forms in Juang and Sora. This category then gradually expanded to include those neighboring categories in Kemmer's (1993: 202) semantic map which were semantically closest to this category—indirect reflexives, the “passive middle” and the passive (cf. once again Figure 6.1).

Finally, any TAM/PERSON-syntagma marked for the passive/reflexive obligatorily appears in the middle voice.

6.5.2 “Durativity”—*continuative* kan, *semel-iterative* lo? and *iterative* khor

There are three markers which may all be subsumed under the general heading “durativity” Although there are subtle differences between them, they are largely interchangeable and it is often difficult to distinguish between them semantically.

In two of the following three categories, the *semel-iterative* and the *iterative*, the semantic base may be a masdar or, less commonly, an underived contentive morpheme. When the base is a masdar and the TAM/PERSON-syntagma is marked by either of these two categories, the TAM/PERSON-syntagma appears in the same basic voice as it would in

that for Pinnow, the reflexive interpretation is secondary to that of the passive, being a kind of semantic side-effect.

the unmarked form, i.e., in the same basic voice as when the semantic base is an underived contentive morpheme and not a masdar. These are the only TAM/PERSON-syntagmas which may have a masdar as their semantic base but which need not appear in the middle voice, unlike the generic function of the middle (6.4.2.2). Finally, none of these three markers has an independent lexical meaning.

Continuative: kan

This category generally denotes that an action continues for a longer time than expected ("keeps on"). It can also be used to express that the subject continues to carry on with one action while performing another:

154. *ho=kaṛ kayom=ta ro kinbhar=jo joʔ kan=te.*
 that=SG.HUM talk=MID.PRS and courtyard=ADD sweep CONT=ACT.PRS
 'She talks and *keeps sweeping* the courtyard.'

With certain contentive morphemes, this category can also express that an action had been going on for a long time in the past before another action occurred:

155. *ji=aʔ de=na=aʔ seŋ=ga baʔ=te ɖuruŋ*
 1SG=GEN come=INF=GEN before=FOC rice=OBL pound
kan=oʔ=ki ghaʔɖ buŋ bhaʔ=ki.
 CONT=ACT.PST=PL therefore finish SUD=MID.PST
 'Before I arrived *they had been pounding* the rice *for a long time*, therefore it (= the rice) was quickly finished.'

Finally, with some contentive morphemes denoting a change-of-state, particularly those of motion but also transitives such as *ɖoʔɖ* 'take (away)', the notion of "continuation" denotes that the change is more-or-less permanent:

156. *ji am=bar=aʔ pothi=te ɖoʔɖ kan=oʔj.*
 1SG 2=2HON=GEN book=OBL take CONT=ACT.PST.1SG
 'I *took* your book (*and I'm not giving it back!*).'

Marked for a person other than the first person, this could also mean that the person who took the book lives far away and will not be able to bring it back for some time.

Semel-iterative: lo?

This category is used primarily by the older generation. For those speakers who do use this form, it marks an event which occurs over and over, but (generally) only on one occasion, in contradistinction to *khor* (see below), although these two are often interchangeable.

With *lo?* the semantic base is normally the masdar. However, the use of the underived monosyllabic form instead of the reduplicated form is not incorrect, although there seem to be slight semantic differences which require further study.

157. *beroʔd loʔ=ki=may ro doko loʔ=ki=may.*
 stand.up S:ITER=MID.PST=3PL and sit.DOWN S:ITER=MID.PST=3PL
 ‘They kept on standing up and sitting down again (e.g., at a meeting).’

There is a very strong association between iterativity and conativity, i.e., an unsuccessful attempt to perform an action, as the following examples illustrate.³⁷

158. *kāṛayboʔ=ki būt=te akeʔd lokh=oʔ=ki muda*
 old.woman=PL gram=OBL chew S:ITER=ACT.PST=PL but
akeʔd=na umay pal=oʔ.
 chew=INF NEG.3PL be.able=ACT.PST
 ‘The old women were chewing the gram over and over but couldn’t chew it (e.g., they tried but it was too tough).’

159. *ho=kaṛ=te moʔd qaʔ dhoʔ=sikh=oʔ ghaʔd chīkay*
 that=SG.HUM=OBL eye water grab=PERF=ACT.PST therefore sneeze
lokh=oʔ muda chīk um muʔ=ki.
 S:ITER=ACT.PST but sneezing NEG emerge=MID.PST
 ‘His eyes filled with water (= eye-water grabbed him), so he kept “sneezing” over and over but no sneeze emerged (i.e., he kept starting / trying to sneeze, but it never came to a real sneeze).’

³⁷ “Pure” conativity is marked by *dakha / lakha* (6.5.8), which may be related to *lo?* etymologically. Recall also from 2.4 above that a morpheme-final /ʔ/ becomes /kh/ before the active past marker =oʔ.

With some contentive morphemes, *lo?* does not imply iterativity but rather a prolonged event. Here, the difference to the continuative maker *kan*, discussed above, is very subtle, if there is one at all:

160. *ijn hoqom kamu=na la?=kho'j. jur-jur ba? kosor*
 1SG other work=INF IPFV=PSIII.1SG spread.out-RDP rice dry

lo?=ki.

S:ITER=MID.PST

'I was doing some other work. The rice which had been spread out [to dry] *was slowly drying.*'

Iterative: khor

Unlike *lo?*, *khor* typically denotes iterativity with an action which comes to a complete stop and then begins again. Here as well, the masdar generally serves as the semantic base, although a non-reduplicated monosyllabic contentive morpheme may also be used.

161. *qubhni=te pe? yo=kon kowa? han=ti'j u=ti'j lej-lej*
 cup=OBL rice see=SEQ crow that=side this=side fly-RDP

khor=ki muda solo?=te yo=kon pe? jo?=na um

ITER=MID.PST but dog=OBL see=SEQ rice eat=INF NEG

are=ki.

descend=MID.PST

'Having seen rice in the cup, the crow *kept flying back and forth*, but having seen the dog, it didn't come down to feed.'

According to our data, the use of *lej-lej lo?=ki* here instead of *lej-lej khor=ki* would denote that the crow kept flying back and forth without landing, whereas the use of *khor* suggests that the crow landed here and there and then flew again. The following presents a further, similar example.

- 162 *kole?=ya? ma=qom ro apa=qom ho=kiyar=te jal*
 parrot=GEN mother=3POSS and father=3POSS that=DU=OBL net

tay oj=na thon han=te u=te

ABL take.out=INF PURP that(DIST)=OBL(= 'there') this=OBL(= 'here')

pu²d-pu²d khor=ta²j=kiyar ro sob=ga “tãy-tãy”
 jump-RDP ITER=MID.PROG=3DU and all=FOC “tay, tay!”

toro²d=taj=ki.

CTY=MID.PROG=PL

‘The [two] parrots’ mother and father *keep jumping here and there* to free them from the net and all [the birds] are crying “*Tay! Tay!*” ’

[BB, 2:34]

The notion ‘here and there’ is quite common with *khor*, primarily with morphemes denoting motion (163) but not exclusively with such morphemes (164).

163. *u bo²=ki ti²j=ga ciŋna=ki buli khor=na*
 this place=PL direction=FOC chick=PL wander ITER=INF

la²=ki=may.

IPFV=MID.PST=3PL

‘The chicks *were running here and there in all directions.*’

164. *no²-no² khor=o².*

eat-RDP ITER=ACT.PST

‘[S/he] ate a little of this and then a little of that (i.e., here and there from the plate)’

6.5.3 *Anticipatory telic dʊ²d and culminatory telic go²d*

These two v2s are largely synonymous and serve to mark an action or change of state as reaching completion. The difference between the two is one of narrative structure.

- *dʊ²d* (with the intervocalic allomorph *dʊt*) indicates that another event follows directly upon the event denoted by the semantic base of the TAM/PERSON-syntagma that it marks;³⁸

³⁸ In fact, *dʊ²d* has virtually the same function in Kharia as the “pluperfect” and “irrealis” markers in *-le* (and variants) in Santali (cf. Neukom, 2001: 79ff.). Especially interesting in this respect is that *dʊ²d* is homophonous with the lexical morpheme ‘take’, from which it undoubtedly derives. This could mean that the form *-le* in Santali derives from Indo-Aryan, as many Indo-Aryan languages have a verb *le-li-* meaning ‘take’, which generally also functions as a v2 in complex predicates. Cf. also Sadri *le-* ‘take’ and *-le* ‘A.TEL’ (Peterson, forthcoming, a); although these two forms are not related to one another in Sadri, they are nevertheless homophonous. *go²d*, on the other hand, does not have an independent lexical meaning.

- *goʹd* (with the intervocalic allomorph *goɾ*) denotes that the event is the final event in a particular narrative sequence. That is, *goʹd* marks a turning point or culmination in a narrative.

To get an idea of the difference between these two categories, consider their use with the morpheme *aw* ‘stay, live’: *aw dʹoʹd=ki=may* [stay A:TEL=MID.PST=3PL], with the anticipatory telic marker, would mean ‘they stayed there (and then . . .)’, implying that their stay at a particular place was short, as they were on their way somewhere else. It is often directly followed by *ro* ‘and’ or *tay* ‘then’ and another TAM/PERSON-syntagma. *aw goʹd=ki=may*, on the other hand, with the culminatory telic marker, would mean that they settled down somewhere to live, and the narrative would then continue on a different topic. The following presents a few examples from texts.

165. *tay raja jhari mudh kalo=ki=te ro yahudi*
 then king all chief priest=PL=OBL and Jew

jait=aʹ etoydaq=aʹ guru=ki=te dʹokloʹ
 ethnic.group=GEN receiving.orders=GEN teacher=PL=OBL meeting

remaʹ dʹoʹh=oʹ ro ho=ki=te “masih=te ate
 call A:TEL=ACT.PST and that=PL=OBL Messiah=OBL where

jorme=na ayiʹjʹ?” gam=oʹ ro juʹj=oʹ.
 be.boim=INF QUAL.PRS say=ACT.PST and ask=ACT.PST

‘The the king *called* all the chief priests and teachers of the Jewish people who received orders [from the king] to a meeting *and* asked (= said and asked) them “Where is the Messiah to be born?”’

[Kullū, 1992: 2 = Matthew, 2:4]³⁹

166. *de babu, amoʹd gujuʹj=na roʹ daʹ=ki uʹd*
 well boy wash.face wash.feet=MID.IRR FOC water=PL drink

dʹoʹɾ=e=m tay dukham sukham=na=pe.
 A:TEL=ACT.IRR=2SG then chat=MID.IRR=2PL

‘Well then, boy, wash your face and feet. Then *you will drink* water (etc., or whatever, = ‘PL’) [*and*] *then* you will all have a chat.’

[Kerkettā, 1990: 26]

³⁹ Translation in the Saint Joseph Edition of *The New American Bible*: “Summoning all of the chief priests and scribes of the people, he inquired of them where the Messiah was to be born.” (Catholic Biblical Association of America, 1970: The New Testament, 6).

The following presents two further examples. Note that a TAM/PERSON-syntagma marked by do^2d need not be followed by another TAM/PERSON-syntagma, as long as it is clear that another action or event is to follow:

167. *gomke, am=ga el=a? gomke heke=m, je=ghay*
 lord 2SG=FOC 1PL.EXCL=GEN lord QUAL.PRS=2SG CREL=way
am karay=na lam=te=m, karay[=e], magar sey
 2SG do=INF want=ACT.PRS=2SG do=ACT.IRR but first
ji=a? kayom se ondor dot=e!
 1SG=GEN word please hear A:TEL=ACT.IRR
 ‘Lord, you alone are our lord, do as you wish to do, but first please
listen to my words!’ [HJPa:49–50]

168. *ondor dot=e=pe sey*
 hear A:TEL=ACT.IRR=2PL first
 ‘First hear this. ’ (i.e., before you continue) [Kereket[ta, 1990: 4]

go^2d is much more common than do^2d and signals not only that the action has reached its conclusion, but also that the story is to continue in a different direction. E.g., in the next example, the Kharia priest has dalled too long and another person, a Brahman, has taken his place and is performing the sacrifice. The Kharia priest then finally arrives and sees this. The text reads:

169. *yo=te la? pujapath=ko suru go^2d=si?, absi?b*
 see=ACT.PRS then sacrifice=CNTR begin C:TEL=PERF begin
go^2d=si?=may.
 C:TEL=PERF=3PL
 ‘He sees, then, that the sacrifice *has* already *begun*, [i.e., they] *have*
 already *started*.’ [AK, 2:29]

The story then continues: Seeing that he had been replaced, the Kharia priest then gives up his priesthood entirely and even begins carrying the new priest around in the palanquin.

6.5.4 Totality may

may denotes the totality of an action or change of state with respect to a particular participant in the scenario. E.g., in the following example this is with respect to the tea, the object:

170. *sori sori caha u¹d may=oʔ=ki (*ro tam jou*
 together REP tea drink TOTAL=ACT.PST=PL and now up.to
u²d=te²j=ki).
 drink=ACT.PROG=PL
 'They *drank up* all the tea together (*and are still drinking).'

Even with transitive activity predicates, the affected entity need not be an object but may also be a plural subject. Here, the clause whose predicate is marked by *may* has a gradual interpretation, i.e., 'one-by-one' until all have performed the action:

171. *ho=kaʔ=te (paʔi paʔi / muɖu muɖu / *sori sori)*
 that=SG.HUM=OBL turn REP one.HUM REP together REP
gil may=oʔ=ki.
 hit TOTAL=ACT.PST=PL
 'They *all hit* him (in turn / one after the other / *together).'

With predicates denoting a simple change of state or a telic activity, the affected entity is usually either an intransitive subject or an object. If the affected entity is plural, *may* denotes that all members of the group have been affected. If the affected entity is incremental or decremental, the TAM/PERSON-syntagma marked by *may* then has a scalar interpretation, i.e., 'really, very, completely'

172. *ka²bʔo sango³d ɖom may=ki.*
 door close PASS TOTAL=MID.PST
 'The door *was shut completely*.'

The affected entity may also be a locative adjunct:

173. *ho=kaʔ sou²b palaŋ=te gitaʔ may=ki.*
 that=SG.HUM all bed=OBL lie TOTAL=MID.PST
 'He *took up the whole* bed (e.g. by tossing and turning).'

When *may* occurs in the past imperfective (6.8.1.1), a conative interpretation generally results.

174. *kahani=te* likha may=na laʔ=ki=may (*muda umay*
 story=OBL write TOTAL=INF IPFV=MID.PST=3PL but NEG.3PL
pal=oʔ).
 be.able=ACT.PST
 ‘They were trying to finish writing the story (but couldn’t).’

For further discussion, see Peterson (in press, b).

6.5.5 Ambulative *san* ‘while going’

This v2 derives from an older contentive morpheme **san* ‘go’ (cf. Santali *sen* ‘go’ and Kharia *sangoʔd* ‘walk’ (< **san* and the culminating telic marker *goʔd* (6.5.3)). It denotes that the TAM/PERSON-syntagma it marks refers to an action that took place while in motion.

175. *ol=na* *olna* *kayar* *beloŋ* *san=ki*.
 bring=INF REP mango ripen AMB=MID.PST
 ‘While being brought [to market], the mangoes ripened along the way.’

With *ol* ‘bring’, *doʔd* ‘take’, *col* ‘go’ and *del* ‘come’, *san* denotes commutativity and is best translated as ‘along with’ (Kullū, 1981: 49). Note that Kullū also writes that *san* is only compatible with these four contentive morphemes and in fact, some speakers we questioned rejected example (175) as ungrammatical.

176. *baysali* *tay* ... *ho=ki* *moŋ* *baysali* *koŋtaŋ* *ro* *baysali*
 Baisali ABL that=PL one Baisali cow and Baisali
sã *ol* *san=oʔ=ki*.
 ox bring AMB=ACT.PST=3PL
 ‘From Baisali. they brought one Baisali cow and Baisali ox along with them.’ [MT, 120]

6.5.6 Suddenness *bhaʔ*, *hambaʔ* and *dhaʔb*

These three markers denote that an event occurred suddenly or quickly. None of the three have an independent lexical meaning. *bhaʔ* and *hambaʔ* would seem to be related to one another etymologically, although if so, the relation between them is unclear.

177. *kiʔjtaʔ=te col bhaʔ=ki u=tay hontayʔ*
 what.time=OBL go SUD=MID.PST here=ABL then
 'When *did [he] suddenly leave* from here, then?'

[Kerkeṭṭā, 1990: 2]

178. *akil=aʔ kayom no musniṅ a=tiʔj tay=ko ḍel*
 mind=GEN opinion CMPL one.day Q=side ABL=CNTR come
hambaʔ=ki moṅ maha dano ro poʔda=yaʔ jhaṅi
 SUD=MID.PST one big demon and village=GEN all
lebu=ki=te diyoga muḍu muḍu=te joʔ=na
 person=PL=OBL daily one.HUM REP=OBL eat=INF

māre=yoʔ.

begin=ACT.PST

'It is believed (= opinion of the mind) that one day a great demon *suddenly appeared* from somewhere and began to eat all the people of the village, one by one.'

[MT, 1:64]

179. *paʔtopur=te=ga ho=ki ikud=ga kisro=sikh=oʔ=ki*
 Patna=OBL=FOC that=PL much=FOC wealthy=PERF=ACT.PST=PL

hin=aʔ caʔḍom ho=ki rohtaspur hinte ḍoko
 that=GEN for that=PL Rohitasgarh LOC sit.down

dhaʔb=na pal=oʔ=ki

SUD=INF be.able=ACT.PST=PL

'In Patna they had become very wealthy, therefore *they were able to settle (= sit) down quickly* in Rohitasgarh.'

[MT, 1:188]

6.5.7 Departive *tu*

This marker denotes that the subject performed some action and then departed. *tu* cannot serve as the semantic base of a TAM/PERSON-syntagma.⁴⁰

⁴⁰ Cf. Santali *-oʔo* with virtually the same function and which also does not have an independent lexical meaning (Neukom, 2001: 142). Cf. also the form *tu-ḍaʔ* 'carry off (of a river)' (*ḍaʔ* 'water') in 4.6.3, which strongly suggests that this category derives from an earlier contentive morpheme.

180. *maha kulam paṭṭopur tay serj muʔ=ki=may. ho=ki*
 big brother Patna ABL first go.out=mid.pst=3pl that=pl
kuḍaʔb aw=ki=may ho=ki=te u gam tu=yoʔ=ki:“...”
 behind qual=mid.pst=3pl that=pl=oobl this say dpt=act.pst=pl
 ‘The elder brothers set out from Patna first. They *said* to those who
 were behind (*before they left*): [MT, 1:160]

6.5.8 Conative dakha / lakha

Conativity (“try”) is expressed through a number of means in Kharia, one of which is *dakha*. The form *lakha* is also occasionally encountered. This marker may be related to the semel-iterative marker *loʔ*, which can also express conativity. It does not have an independent lexical meaning.

181. *aḍ=ki=yaʔ ompay paro=na boʔ=te=jo mudui=ki*
 ANAPH=pl=gen river cross=inf place=oobl=add enemy=pl
kuday dakha=sikh=oʔ=ki,
 chase conat=perf=act.pst=pl
 ‘The enemy_i also *tried to chase* [them_j away] at the place where
 they_j had crossed the river, [MT, 1:167]

6.5.9 Excessive bay

bay is homophonous with, and undoubtedly derives from, the contentive morpheme *bay* ‘make, do’. As a v2 it serves to denote that the action was “more intense” than the unmarked form or even excessive, somewhat similar to the function served in English by the adverbs ‘extremely’, ‘strongly’ or ‘intensely’.

182. *muʔjḍaʔ pheinga=te leʔ bay=oʔ, “u cimi*
 ant grasshopper=oobl curse exces=act.pst this sugar
ḡḡ=aʔ bohini=na=ḡḡ=ki thuray=siʔ=may.”
 1sg=gen sister=1poss=1sg=pl gather=perf=3pl
 ‘The ant *gave* the grasshopper a *bad scolding*, “This sugar my sisters have gathered.”’ [TK, 2:42]

The compatibility of *bay* with contentive morphemes is highly idiosyncratic and its semantic contribution is often unpredictable. For example, it is compatible with *aʔ* ‘splash’ and *gil* ‘beat’ with an “excessive” or

“intensifying” meaning but also with *kayom* ‘talk’, with the resultant meaning ‘deceive’ On the other hand, it is not compatible with *kamu* ‘work’ or *ge⁷b* ‘burn’, to name just two examples.

6.5.10 *Autopoesis jom*

With avolitional predicates, this marker denotes that something happened on its own, i.e., there was no outside force which caused it to happen. With potentially volitional predicates on the other hand, it denotes that the agent simply performed the action because s/he wanted to and was under no obligation to do so. It is best translated by the English word *just*. It is not used as an independent lexical morpheme, although, as Pinnow (1966: 112–3) notes, it derives from an earlier contentive morpheme meaning ‘eat’

183. (**bero bun*) belom jom=ki.

SUN INST ripen AUTOPOES=MID.PST

‘It (e.g., fruit) *just ripened* (i.e., *all by itself*) (*through / *because of the sun).’

184. *in kothri=te badli* jom=o⁷j.

1SG room=OBL change AUTOPOES=ACT.PST.1SG

‘I just *changed* the room *around* (for no particular reason).’

The use of this category can also have an impolite connotation, as in the following example. Here, the person who has sat down has done so simply because he wanted to and it is strongly implied that he did not have permission to do so and had not been offered a seat.

185. *doko jom=ki*.

sit.down AUTOPOES=MID.PST

‘He just sat down (i.e., no one gave him permission).’

6.5.11 *Benefactive kay*

As the name implies, this category denotes that an action was carried out on behalf of someone other than the subject.

In the following example from a children’s story, a boy is in a tree in which loaves of bread grow. An elderly woman (actually, a witch) is waiting at the bottom of the tree and, thinking she is waiting for something to eat, he throws her down two loaves.

186. *hobne=te=ga kongher kãraybo?=te yo=yo? ro*
 that.much=OBL=FOC boy old.woman=OBL see=ACT.PST and
ubar koloŋ ho=kaŋ=a? mugamte ob-gur
 two bread that=SG.HUM=GEN in.front.of CAUS-fall
 kay=o?

BEN=ACT.PST

‘After a while, the boy saw the old woman and *threw down* (= *caused to fall*) in front of her two loaves of bread *for her*.’ [BB, 1:40]

kay is homophonous with the contentive morpheme *kay* ‘lift’, although it seems unlikely that it has derived from this form.

6.5.12 *Seldomly encountered v2s*

The following examples are given with little comment as the respective *v2s* are either unique or at least very rare in our data and require further study.

6.5.12.1 col ‘go’

This morpheme, when serving as a *v2*, appears to denote durativity of some type, in this example a continuing, long-term event.

187. *u=ghay kam=ga hoy=ga col=ki=may.*
 this=way less=FOC become=FOC *v2:go=MDD.PST=3PL*
 ‘Thus, they became fewer and fewer [= *went becoming* few].’
 [MS, 1:173]

6.5.12.2 oʃj ‘take out’

All that can be said about this *v2* at the moment is that it appears to denote telicity. There are no examples of this in our corpus, only in Pinnow’s texts.

188. *ro kaiko hajar memon=te nuh=te ray*
 and several thousand year=OBL Noah=OBL chose
och=o?, ro nawka bay=na etoŋ=o?, ...
v2:take.out=ACT.PST and ship make=INF order=ACT.PST
 and after (= in) several thousand years He (= God) *chose*
 Noah, and ordered [him] to build a ship, [HJPa:119, ln. 6]

A very similar example is given below, in which the semantic base of the TAM/PERSON-syntagma is marked for the sequential converb, which here separates the semantic base from the v2. This construction is typical of many Indo-Aryan languages, such as Nepali (cf. Peterson, 2002b), but is highly atypical of Kharia.

189. *ho=kaṛ moṅ upay socay=kon oḥ=o?*, ...
 3=SG.HUM one means think=SEQ take.out=ACT.PST
 'He thought up a means . [HJPa:38]

6.5.12.3 *ol* 'take'

As a v2, *ol* would seem to be a telicizer and its use is very rare (only two examples, both from the same speaker of the southern dialect). Note also that the v2 in both examples is separated from the semantic base of the TAM/PERSON-syntagma by the focus marker =*ga*, an example of which is given here.

190. *la? ho=te khali khariya bheir aw=ki=may*
 then that=OBL(= 'there') only Kharia much live=MID.PST=3PL
oḍo? hoḍom jait=ko mane ho=ki kuday=ga
 and other ethnic.group=CNTR umh that=PL chase=FOC
ol=te'j=ki.
 V2:bring=ACT.PROG=PL
 'Then only many Kharia lived there and, umh, they *are driving away* the other ethnic groups.' [MS, 1:153]

6.5.12.4 *pal* 'finish'

As a v2, *pal* denotes that an action reaches completion.

191. *chiṛopoṛo kolon laḍ=na maṛe=yo?=ki. laḍ*
 quickly bread bake=INF begin=ACT.PST=PL bake
pal=ke cah sobre=yo?=ki.
 v2:finish=SEQ tea prepare=ACT.PST=PL
 they quickly began to bake bread. When they *had finished baking*, they prepared tea.' [HJPa:76]

pal is also found as a contentive morpheme with a very similar meaning. Cf. the following, from the same page as the previous example:

192. *bero dharke=na jaw khorī=ya? jharī=ki=ga ađi*
 sun go.down=INF up.to village=GEN all=PL=FOC ANAPH
ađi=ya? kamu=ki=te pal=o?=ki.
 REP=GEN work=PL=OBL finish=ACT.PST=PL
 ‘By the time (= until) the sun went down, all of the villagers *finished*
 their work.’ [HJPa:76]

The v2 *pal* does not appear to have ever been very common—only two examples of it were found, both in the first half of the texts in Pinnow (1965a). The only similar construction in our own data, also seldom, separates the semantic base of the TAM/PERSON-syntagma from the v2 by the sequential converb marker, similar to *oʹj* in 6.5.12.2.

193. *muda je kamu=na laʹ=ki ho=je? uʹđ*
 but CREL work=INF IPFV=MID.PST that=SG.NHUM drink
joʹ=kon=ga pal=na laʹ=ki.
 eat=SEQ=FOC finish=INF IPFV=MID.PST
 ‘But whatever he did, it was just enough to live on (= he *used to finish it, having eaten and drunken*).’ [RD, 2:36]

The v2 *pal* ‘finish’ is homophonous with the “auxiliary” *pal* ‘be able’ (6.8.2.3). The two differ morphosyntactically in that the abilitative *pal* requires an infinitive whereas the telic *pal* follows the semantic base (optionally marked by the sequential converbal marker).⁴¹

6.5.12.5 *saphay* ‘clean’

There are two examples of *saphay* functioning as a v2 in our corpus, both from the same speaker of the southern dialect, and apparently none in Pinnow’s texts. Due to the paucity of examples, the function of this v2 is somewhat unclear, but note that in both examples it is found in conjunction with a semantic base denoting a negative action or feeling.

⁴¹ Similarly, in Nepali the v2 *sak* ‘finish’ is homophonous with the auxiliary *sak* ‘be able’ Otherwise, this polysemy appears to be rare in the region.

194. *khariya a=te pal=o?=ki hin=te=ga*
 Kharia Q=OBL be.able=ACT.PST=PL that(DIST)=OBL(= 'there')=FOC
kackac saphay=o?=ki.
 oppress V2:clean=ACT.PST=PL
 'Wherever they could, *they oppressed* the Kharia.' [MT, 1:18]
195. *po?da=? jhari lebu=ki botoŋ saphay=ki=may.*
 village=GEN all person=PL fear V2:clean=MID.PST=3PL
 'All the people of the village *were very afraid.*' [MT, 1:65]

6.6 PARTIALLY FINITE AND NON-FINITE FORMS

TAM/PERSON-syntagmas in Kharia which are not fully finite may either be marked as explicitly non-finite, e.g., through an infinitive or converb marker, or they may be partially finite, lacking some of the marking necessary to be able to function as a matrix predicate. Fully finite TAM/PERSON-syntagmas are marked for PERS/NUM/HON, whereas non-finite and partially finite forms lack this marking.

6.6.1 *Partially finite forms*⁴²

Partially finite forms are similar to TAM/PERSON-syntagmas whose semantic base consists of more than one sub-base (generally a simple contentive morpheme, cf. 6.3.2) but differ slightly from these. Recall that the various sub-bases of a TAM/PERSON-syntagma with multiple sub-bases are directly juxtaposed to one another with no intervening marking, as in (196).

196. *tay adj=te=jo batay=o? no "dhãy[=e] am, han*
 then ANAPH=OBL=ADD tell=ACT.PST Cmpl hurry=ACT.IRR 2sg that
maŋa bo?te qa[?] [a]yi?j. uđ de=na=m."
 cave LOC water QUAL.PRS drink come=MID.IRR=2SG
 'Then [he.] said to him, too, "You hurry, at the cave there is water.
Drink [and] come [back]" ' [AK, 1:38]

⁴² The discussion here is largely based on that in Peterson (2002a) but also incorporates much new data.

Partially finite forms are formed similarly, but with these TAM/PERSON-syntagmas each non-final sub-base contains a certain amount of inflectional marking without being fully finite, as at least one marker from the functional head of the predicate is missing. In these non-final sub-bases, person marking is always lacking, as with *su=yo?* in (197), although other marking may also be lacking (e.g., (202) below). The two sub-bases may or may not be joined by a conjunction such as *ro* or *oɔo?*, both of which mean ‘and’, and they may also be separated by arguments or adjuncts of the following sub-predicate (e.g. (202)).

197. *kono no maha jhari=ga tomne tomne lutui ikud*
 small or big all=FOC new REP clothes very
kelom-bo? su=yo? aŋku=yo?=ki.
 beautiful-INTENS put.ON=ACT.PST wear.a.chaddar=ACT.PST=3PL
 ‘Whether big or small, all *put on* [and] *wrapped around* [themselves] the very beautiful new clothes.’ [Pinnow, 1965: 75]

Although the data at the moment do not allow a full analysis of these junctures in terms of their nexus relations, it seems safe to say that they are core junctures as the two or more sub-bases obligatorily share the same subject, and are probably of the cosubordination type in that they share all clause- and core-level operators, especially tense and modality (cf. Foley & Van Valin, 1985; Van Valin & LaPolla, 1997). This awaits further verification.

Partial finiteness is highly restricted and is typical of certain fixed expressions and events consisting of sub-events which are considered to be highly related. It is also much more common in the past tense, where it is more-or-less obligatory for certain combinations, than in, e.g., the irrealis, although it is possible in all categories. Also, at least in the modern language, is restricted to the active voice in those categories which have the active/middle distinction.⁴³

⁴³ Partial finiteness would seem to go back to Proto-Munda, as similar data are also found in North Munda languages, e.g., Santali (Neukom, 2001: 176f).

See also under the discussion of root types in Remo (Fernandez, 1967: 37) under §2., *Root + o? + Root*, where *-o?* is homophonous with—if not identical to—the marker of the active past, which Fernandez refers to as Class I, past. It is perhaps significant that the past active form appears to have been generalized in Remo and that this construction in Kharia is much more common in the past and does not occur in the middle voice. Nevertheless, Fernandez differentiates between forms with complex roots of the form “Root + /o?/ + Root” and “Root + /o?/ + Root + /o?/”, which shows that the situation in Remo

Partial finiteness is especially common with certain everyday actions which tend to be done together, such as eating and drinking. Consider the following example, adapted from Roy & Roy (1937: 180f.).

198. *u=kiyar taŋ el=a? sori=ga ɲog=e*
 this=DU now 1PL.EXCL=GEN with=FOC eat=ACT.IRR

uɖ=e=kiyar.

drink=ACT.IRR=DU

'They two *will now eat and drink* with us.'

The use of partially finite sub-predicates in the irrealis is especially common in the texts in Pinnow (1965a) in contexts in which what is to be done in a ceremony is being described.

199. *.beʔa=? cunɖu aŋri=te sundrom lebʔo=ye ro*
 boy=GEN index.finger finger=OBL vermilion smear=ACT.IRR and

beʔi=ya? maŋ=te joɖ=e=ki, . . .

girl=GEN hair.parting=OBL wipe=ACT.IRR=PL

' *they should / will smear* vermilion on the index finger of the boy *and wipe* [it] on the parting of the girl's hair,

[HJPa:155,23]

200. *enem koŋ-kon buŋ acaka ɖel goɖ=na=ki la?*
 without know-RDP INST sudden come C:TEL=Mid.IRR=PL then

baherte goŋ=e ɲog=e=ki.

outside cook=ACT.IRR eat=ACT.IRR=PL

'If, without knowing [that the family is impure], they suddenly come, then *they will cook and eat* out-of-doors.' [HJPa:144-5]

is somewhat different from Kharia, where an intervening =oʔ is only found when the entire predicate is marked for active, past. In other words, both sub-predicates in Kharia must have the same TAM-values, whereas in Remo, an intervening -oʔ is also found when the entire predicate is marked for some other TAM-category, although Fernandez gives no examples. Most interestingly, Fernandez notes that he has only one example of the type "Root /oʔ/ + Root + /oʔ/", *pug-oʔ-ruŋ-oʔ* 'cut (*pug*) and bring (*ruŋ*)', which makes it clear that in Remo the entire predicate is generally NOT marked for -oʔ but rather only the first sub-predicate.

Partially finite sub-predicates are always marked for (at least) one element less than the sub-predicate which they are morphologically dependent on. In the preceding examples this meant that the non-final sub-element was not marked for PERS/NUM/HON, while the final element was. Now consider the following example:

201. *beʃi tij=aʔ lebu beʃa=te paŋ god*
 girl side=GEN person boy=OBL carry.on.shoulder C:TEL
ʒi<ʔbh>ar=e ro ʒoko=ye=ki beʃa
 enter-<CAUS>=ACT.IRR and sit.down-<CAUS>=ACT.IRR=PL boy
beʃi=te panti=te, ...
 girl=OBL row(n.)=OBL
 ‘The people from the girl’s family (= side) *will carry* the boy on their shoulders into [the house] (= *will enter*) *and set* them down next to one another (= in a row), [HJPa:155, 20]

The predicate here actually consists of three sub-predicates: The first element, *paŋ god*, is partially finite and contains less “finite” marking than the following sub-predicate, *ʒi<ʔbh>ar=e*. It is thus morphologically marked as being dependent on this latter element by lacking a TAM-element, in this case the irrealis active marker =*e*. These two together form a unit, both semantically and morphologically speaking. The first sub-predicate modifies the second, describing how the boy is to be brought in—and the two together compose the first (complex) sub-event of the event. This complex unit is then morphologically marked as dependent on the final sub-predicate, *ʒoko=ye=ki*, which is fully finite, through the lack of person marking on *ʒi<ʔbh>ar=e*. The following presents another example of this type of partial finiteness:

202. *am=aʔ moʒ=te japiʒ oʒoʔ tanij deri=yaʔ ghaʒ*
 2SG=GEN eye=OBL close.eyes and a.little time=GEN for
socay=e
 think=ACT.IRR
 ‘Close your eyes and *think* for a moment.’ [HJPa:95]

As the final sub-predicate in (202) is not marked for person, being a second-person singular imperative (6.4.1.4), the only way the non-final

sub-predicate can be marked for “less finiteness” is by omitting the marker for the irrealis active on *japīd*.

Despite the fact that all speakers we consulted insisted that partially finite non-final sub-predicates appear in the active voice where an active/middle distinction is found, there are also counter-examples to this in Pinnow’s texts, although such examples are seldom. The following presents two such examples. Examples such as these were rejected by speakers we consulted, hence this either held for Kharia in general at an earlier time or the few examples found in Pinnow’s texts are restricted to certain dialects or perhaps idiolects.

203. *ol=e* *ro* *ḍoko=na=ki*
 bring=ACT.IRR and sit.down=MID.IRR=PL
 ‘they will bring [it] and sit down’ [HJPa:161,6]

204. *onḍor=e* *ro* *sikhe=na=niḡ*
 hear=ACT.IRR and learn=MID.IRR=1PL.INCL
 ‘we should hear and learn [His teachings]’ [HJPa:91]

6.6.2 *Non-finite forms*

6.6.2.1 *The masdar*⁴⁴

Every (fully or partially) finite TAM/PERSON-syntagma whose semantic base consists of a simple contentive morpheme (possibly also marked for derivational categories such as the causative or reciprocal) has a corresponding form which will be referred to here as the masdar. It lacks all TAM and person marking (except for certain v2s, see below), although it retains valency-related marking such as causative and passive / reflexive marking. It may govern an object and adjuncts, which then take the same marking they would have with a finite TAM/PERSON-syntagma, but what corresponds to the subject of the finite TAM/PERSON-syntagma appears in the genitive with the masdar if it is expressed. The masdar may also mark for case.

The masdar ultimately derives from the bisyllabic constraint on phonological words. Its primary function is that of secondary or “non-finite” predication, but it may also be used referentially and even in primary or “finite” predicative function, in addition to its attributive use. As this

⁴⁴ Cf. also the discussion of this form with respect to parts of speech in 4.3.2.

form—like all semantic bases in the language—is not specified for any particular discourse-pragmatic function, it will be referred to here as the “masdar”, a term borrowed from Arabic linguistics, where the form this term refers to there is not a verbal noun, as it is usually translated, but rather “a form which is underspecified with respect to its syntactic potential, one which can be both a complement to the predicate, i.e., it can assume nominal functions, and can be a predicate. One trait which it shares with the verbal noun in Indo-European languages is that it can mark for case.” (Maas, 2007: 134, this author’s translation).

The masdar is formed as follows: Polysyllabic semantic bases undergo no changes, while monosyllabic forms reduplicate. This is thus a purely phonological process. “Semantic base” here includes at least the underlying contentive morpheme, causative marking, the passive / reflexive v2 *ɖom* and the benefactive v2 *kay*. The following provides a few examples:

205. <i>Underlying form of the semantic base</i>	<i>Masdar</i>
<i>aw</i> ‘live, stay, remain’	<i>aw-aw</i>
<i>bay</i> ‘make’	<i>bay-bay</i>
Cf. <i>bay kay</i> [make BEN]	<i>bay kay</i>
<i>borol</i> ‘live’	<i>borol</i>
<i>kersonj</i> ‘marry’	<i>kersonj</i>
<i>likha</i> ‘write’	<i>likha</i>
<i>osel</i> ‘become white’	<i>osel</i>
<i>ruʔ</i> ‘open’	<i>ruʔ-ruʔ</i>
<i>rusuŋ</i> ‘become red’	<i>rusuŋ</i>
<i>soŋ</i> ‘buy’	<i>soŋ-soŋ</i>
Cf. <i>ob-soŋ</i> ‘sell’ [CAUS-buy]	<i>ob-soŋ</i>
<i>yo</i> ‘see’	<i>yo-yo</i>

There are no data at our disposal on whether other v2s or the reciprocal marker *kol* also belong to the semantic base, but the TELIC v2s such as *goʔɖ* and *ɖoʔɖ* or the perfect marker *siʔ* clearly do not belong to the semantic base, as the following examples show.

206. <i>ɟɪ=aʔ</i>	<i>bay-bay</i>	<i>oʔ</i>	‘the house I built’
1SG=GEN	make-RDP	house	

207. *jn=a?* bay kay o? 'the house I built (for someone)'
1SG=GEN make BEN house
208. **jn=a?* bay=si? o? 'the house I have built'
1SG=GEN make=PERF house
209. **jn=a?* bay go³d o? 'the house I built'
1SG=GEN make C:TEL house

As noted, what corresponds to the subject of a finite TAM/PERSON-syntagma appears in the genitive with a masdar, if it is expressed:

210. *no prabhu=wa? janam=te mane ho=ki=te iswar=a?*
Cmpl God=GEN birth=OBL umh that=PL=OBL Lord=GEN
ḍaṅ-ḍaṅ aw=ki, . . .
send-RDP QUAL=MID.PST
that at God's birth, umh, the Lord had sent to them [His Son]
(= [Jesus] was a *sent-one of the Lord*), [MS, 1:289]
211. *la?²=ko* *darhi=ya?* *ropa* *jal=te* *bajhe*
then=CNTR (= 'but') Darhi=GEN spread.out net=OBL get.trapped
go²d=na *la?²=ki=may.*
C:TEL=INF IPFV=MID.PST=3PL
'But they would get trapped in the net *that Darhi had spread out.*'
[BB, 2:11]
212. *ro anaj=te=jo* *purkha=ki=ya?* *ter-ter* *kamu=te=ga*
and 1DU.INCL=OBL=ADD ancestor=PL=GEN give-RDP work=OBL=FOC
karay=kon borol=na ayi?².
do=SEQ live=INF QUAL.PRS
'And we both must also live by doing the work *which the ancestors have given us.*'
[BB, 2:46]

As already noted, the masdar is most commonly found in attributive function, in what translates as a relative clause. The following presents a few further examples:

213. *jal=te bajhe konthed=ki=te yo=ta yota darhi lere?*
 net=OBL get.trapped bird=PL=OBL see=CVB REP Darhi joy

jal tay konthed=ki=te oj=kon ađi=ya? bora
 net ABL bird=PL=OBL take.out=SEQ ANAPH=GEN sack

theila=te sajay=na la?=ki.
 (hand)bag=OBL put.into=INF IPFV=MID.PST

‘Seeing the birds *caught in the net*, Darhi would joyously take the birds out of the net and put them into sacks and bags.’ [BB, 2:13]

214. *iji musa tay=ga konthed bajhay kamu melay=kon*
 1SG today ABL=FOC bird trap work leave=SEQ

hođom kamu kara[y=e]=ij.
 other work do=ACT.IRR=1SG

‘As of today, I will give up the work *of trapping birds* and do other work.’ [BB, 2:50]

The masdar may also be used without an explicitly mentioned lexical head if the identity of this entity is clear from context:

215. *am=ga patar ter-ter heke=m, am, yesu,*
 2SG=FOC light give-RDP QUAL.PRS=2SG 2SG Jesus

đe=na=m ro patar ter=e=m patar
 come=MID.IRR=2SG and light give=ACT.IRR=2SG light

ter=e=m.
 give=ACT.IRR=2SG

‘You are *the [one] who gives light*, you, Jesus, will come and give light, you will give light.’ [RK, 4:2]

For further examples of the use of the masdar in predicative attribution, see 7.6.2.

Masdars have no inherent orientation with respect to subject, object, instrument, locative, etc., nor with respect to tense.⁴⁵

⁴⁵ Note that *in=a?* in (216) appears in the genitive as it is the “subject” of the masdar, not because of what would appear to be the head nominal (*lebu*), similar to the use of the genitive in example (210) above.

216. *ip=a? yo-yo lebu*
 1SG=GEN see-RDP person
 'the person I saw / see / will see / should see...'
217. *ip=te yo-yo lebu*
 1SG=OBL see-RDP person
 'the person who saw / sees / will see / should see... me'
218. *ip=a? dura=te ru?-ru? kujji*
 1SG=GEN door=OBL open-RDP key
 'the key I opened / open / will open / should open... the door with'

The masdar may also be the lexical head of a Case-syntagma.⁴⁶

219. *o?=ya? bay-bay um=ip ba'j=ta.*
 house=GEN build-RDP NEG=1SG like=MID.PRS
 'I don't like (the act of) building houses.'
220. *ho=kaɾ=a? ter-ter yo=yo'j.*
 that=SG.HUM=GEN give-RDP see=ACT.PST.1SG
 'I saw him give / his (act of) giving [something to someone].'

There is some evidence that the masdar once served some of the functions now served by the infinitive. For example, the masdar of *lam* 'look for', *lam-lam*, is often found in constructions in which one otherwise finds the infinitive, as in purpose clauses.

221. *beɬ=no=m=ki=te kinir qaŋ goɾ=e, lam-lam*
 son=2POSS=2SG=PL=OBL forest send C:TEL=ACT.IRR search-RDP
 co=na=ki.
 go=MID.IRR=3PL
 'Send your sons into the forest, they will go to hunt.' [MS, 2:7]

⁴⁶ The conditions under which a non-subject appears in the genitive are not entirely clear, although the non-subject (of the corresponding finite clause) generally appears in the same case it would appear in in a finite clause. In the case of (219) this would appear to be due to the fact that the oblique-case marking on *o?* 'house', i.e., *o?=te* [house=OBL], preferentially yields a locative interpretation, not that of an object.

222. *hin=a? ghaḍ=ga manu purkha buḥha aḍi=ya? ghol*
 that=GEN PURP=FOC Manu ancestor old.man ANAPH=GEN ten

beṭa=ḍom=ki=te kinir lam-lam ḍaṅ=o?

son=3POSS=PL=OBL forest search-RDP send=ACT.PST

‘For that very reason, the elderly ancestor Manu *sent* his 10 sons into the forest *to hunt*.’ [MT, 1:129]

However, apparently, no other masdar can be used this way: A large number of other masdars were tested in such environments and all were rejected by all speakers consulted, with the exception of *lam-lam*, which was accepted by all speakers. Obviously, the use of *lam-lam* in this construction is not proof that the masdar was once productively used this way, but as such reduplicated forms do serve as infinitives at least in some other South Munda languages, e.g., Remo (Fernandez, 1967: 122), it is certainly a possibility.

Finally, the masdar is also used in a construction to denote a non-changing or even permanent situation / event. In this highly productive construction the masdar appears twice: The first form is marked for the genitive. This is then followed by the masdar marked by the focus marker =*ga*. The functional head of the predicate is then the corresponding finite form of *aw goḍ=* ‘remain C:TEL=’, marked for tense and person.

223. *aba=ḍom rupaya=te yo-yo=ya? yo-yo=ga aw*
 father=3POSS money=OBL see-RDP=GEN see-RDP=FOC remain

goḍ=ki.

C:TEL=ACT.PST

‘His father *just kept staring and staring* at the money.’ [BB, 2:60]

224. *raḥ-raḥ=ya? raḥ-raḥ=ga aw goḍ=ki.*

blossom-RDP=GEN blossom-RDP=FOC remain C:TEL=ACT.PST

‘It just stayed a flower (i.e., it just kept blossoming but never bore fruit).’

6.6.2.2 Infinitives

Kharia is one of the few Munda languages—perhaps the only one—which possess a general infinitive form, marked by =*na*. What corresponds to the subject of a finite TAM/PERSON-syntagma appears in the genitive if it is expressed (225), while objects of an infinitive have the same mark-

ing as with finite TAM/PERSON-syntagmas (226), (227), first example. The infinitive may serve as a Case-syntagma, as in (225), where it is a case-marked adjunct, or in forms such as *noʔ=na* 'food' (literally: '(to) eat'), *uʔd=na* 'drink (n.)', *taʔj=na* 'distribution'. It is also found in modification (cf. (226) and 7.6.2) and is used with a number of auxiliaries to form complex predicates ((227), third example, and 6.8).

225. *am=aʔ jib=na=te ʔhik um hoy=ki.*
 2SG=GEN touch=INF=OBL good NEG become=MID.PST
 'Through your touching [it, it] has become spoiled (= did not become good).' [MS, 1:247]

226. *musniŋ u=ghay hoy=ki no ʔoli=te ʔoʔd=na*
 one.day this=way become=MID.PST Cmpl palanquin=OBL take=INF
bhere u kharīya pahan=te haḍa laʔ=ki.
 time this Kharia priest=OBL pee EMOT=MID.PST
 'One day it happened this way that, **at the time** [they were] to take away the palanquin, the Kharia priest had to pee.' [AK, 2:8]

227. *iḍib=te raksin kaṛayboʔ kongher=te dhoʔ=na=yaʔ upay*
 night=OBL witch old.woman boy=OBL grab=INF=GEN means
konod=na buŋ lemeʔd=na=jo um pal=oʔ.
 think.up=INF INST fall.asleep=INF=ADD NEG be.able=ACT.PST
 'At night, the old witch, because of her thinking up **a means of grabbing the boy, could not fall asleep.**' [BB, 1:51]

Occasionally, it is also found used as an imperative when negated by the Indo-Aryan loan word *na* 'NEG'

228. *laʔ am na cori karay=na, na jahāy=aʔ jan*
 then 2SG NEG theft do=INF NEG INDEF.HUM=GEN life
tar=na, na jahāy=aʔ jahā karay=na.
 kill=INF NEG INDEF.HUM=GEN INDEF.NHUM do=INF
 'So do not *steal*, do not *kill* anyone, do not harm anyone (= do not do anyone's anything).' [MS, 1:309]

At first glance, the form =*na* would appear to have been borrowed from Indo-Aryan (cf. the Hindi infinitival ending *-na*). There is however

another, much more likely candidate: Note that the middle irrealis ending is also =*na* and the infinitive in Kharia may have evolved from this form. An analogy for this interpretation is found in the Tibeto-Burman language Newari, spoken in Nepal, where a nonpast verbal form also functions as an infinitive (cf. Genetti, 1991: 238f and 252, n. 19). This would seem to be a more likely explanation for the emergence of an infinitival marker in Kharia since the infinitive in Sadri, with which Kharia has had a much longer period of contact than with Standard Hindi, is marked by *-ek*, not *-na*.

Finally, it is perhaps also worth mentioning that there is an infinitival form in *-na* in the North Dravidian language Kurukh as well, with which Kharia has also long been in direct contact.

In view of all this, Hindi as the source of the infinitive marker in Kharia appears highly unlikely. If the form has been directly borrowed from another language, this would more likely be Kurukh than Standard Hindi, although a language-internal explanation seems most likely.

6.6.2.3 *Sequential converbs (“conjunctive participles”)*⁴⁷

There are three markers in Kharia which may be considered sequential converb markers: =*ker*, =*ke* and =*kon*. =*kon* is often realized as =*kan* in more southerly dialects. These forms correspond in their functions almost completely to what are traditionally referred to as “conjunctive participles” in South Asian linguistics. The first two are direct borrowings from Sadri (cf. Jordan-Horstmann, 1969: 96f.). Also, similar to the “conjunctive participles” of other South Asian languages, these markers denote, among other things, that the two or more (sub-)predicates are portrayed by the speaker as being directly related to one another in some way, combining to denote a larger, more complex event.⁴⁸ Sequential and other temporally related events which the speaker does not wish to be portrayed as being related are expressed by other means, such as conjunctions and postpositions (7.5.1).

⁴⁷ The status of the converbal marker requires further research. Peterson (2008) treated these markers as separate “words”, as two speakers I had spoken with (hesitantly) accepted forms in which the converbal marker is separated from the semantic base of the predicate by a floating clitic. This suggests that the converbal marker is both a syntactic and a phonological word.

However, as this issue has only been discussed with few speakers, and as these speakers only grudgingly accepted such examples, I will treat these markers here as enclitics and leave the determination of their exact status as “words” to future research.

⁴⁸ Cf. Peterson (2002b) for literature dealing with this topic.

=*kon*, perhaps the more common form, appears to be a calque of the Sadri form *-ker*. Like the cognate form *-kar* in Hindi, *-ker* in Sadri appears to derive from the root *kar-* 'do'. The sequential converbal marker =*kon* in Kharia apparently derives from a similar contentive morpheme, i.e., *ikon* 'make, do'.

The core function of these forms is to denote an action which begins before the eventuality expressed by the finite TAM/PERSON-syntagma.

229. ... *ka?* *kom=ki* *ḍho?=ke* *mu?* *go²ḍ=ki=may*.
 bow arrow=PL grab=SEQ emerge C:TEL=MID.PST=3PL
 they took their bows and arrows and set off (i.e., 'having taken their bows and arrows, they emerged').' [AK, 1:11]

230. *dhirom* *dhirom* *khaṛiya* *maha* *roke²d-lo?* *paro=kon* *rusuṅ*
 slowly REP Kharia big sand-place cross=SEQ red
samuder *ḍam=ki=may*.
 ocean arrive=MID.PST=3PL
 'Slowly but surely, the Kharia *crossed* the great desert and arrived at the Red Sea.' [MT, 1:21]

These converbal markers are also often used to denote the manner in which an action is carried out. In these cases, the action denoted by the converb is generally a more exact specification of the event denoted by the morphologically finite TAM/PERSON-syntagma, as in the following two examples.

231. ... *lay* *koj=kon* *go²juṅ* *bay=si?=may*.
 dig scrape=SEQ path make=PERF=3PL
 they have built the path *by digging and scraping* [the dirt away].' [MT, 1:45]

232. *ho=kaṛ=a?* *purkha=ki=jo* *konthed* *bajhay=na=ya?*
 that=SG.HUM=GEN ancestor=PL=ADD bird trap=INF=GEN
kamu *karay=kon=ga* *borol=na* *la²=ki=may*.
 work work=SEQ=FOC live=INF IPFV=MID.PST=3PL
 'His ancestors as well used to live by trapping birds (= *doing* bird-trapping work).' [BB, 2:7]

Often, however, the sub-predicate marked by the converbal marker simply denotes a simultaneous action which further specifies the finite TAM/PERSON-syntagma:

233. *raksin oqo? jughay khisay=ta ro gone keb=kon*
 witch even more angry=MID.PRS and tooth grind=SEQ
gam=te: . . .
 say=ACT.PRS
 'The witch grows even angrier and . . ., *grinding* her teeth, says .
 [BB, 1:49]

It might be argued that what is being referred to here as the sequential converb is actually a general converb, unspecified for whether it denotes an event which occurs prior to or simultaneous with that of the finite TAM/PERSON-syntagma, and examples such as the last three would clearly seem to favor such an analysis.

However, the sequential interpretation is far more common than the simultaneous interpretation, and a sequential interpretation is almost always possible in addition to the simultaneous interpretation. Thus, the sequential use of this category is clearly the central use, and it would seem that the use of the sequential converb to denote simultaneous events as a kind of general "adjunct" marking has only arisen secondarily.

It is also worth mentioning that this dual usage of "sequential" converbs is typical of many, if not most, South Asian languages.⁴⁹ In fact, it was noted above that not only are the markers =*ker* and =*ke* in Kharia both borrowed from Sadri, the form =*kon* itself would appear to be a calque of the Indo-Aryan construction: The primary function of this category is to express sequentiality, although it is also used to denote "adverbs of manner", just as with the Indo-Aryan construction. This fits in well with the discussion in Malhotra (1982: 217) and Abbi (1993: 549) where it is noted that the use of the sequential converb in Kharia has been borrowed from Indo-Aryan, with both referring to works by Pinnow.

As the examples above show, the converbal clause usually precedes the finite TAM/PERSON-syntagma. Occasionally, it is also found following the TAM/PERSON-syntagma, as in the following example.

⁴⁹ Cf. e.g. Peterson (2002b) for a discussion of Nepali, where the sequential converb has an almost identical range of uses as that of Kharia.

234. *han=tiʔj u=tiʔj buli=na bulina ho=kaṛ rel paṭari*
 that=side this=side wander=INF REP that=SG.HUM train track
tiʔj col=ki laʔ yo=te laʔ=ko soub
 toward go=MID.PST then see=ACT.PRS then=CNTR all
merom=ki goʔj may=ki=may, gaṛi buṅ oton ḍom=ke.
 goat=PL die TOTAL=MID.PST=3PL car INST press PASS=SEQ
 'Wandering this way and that, he went towards the railroad tracks
 and sees that all the goats had died, *having been crushed* by the
 train.' [RD, 1:7]

There is a strong tendency in Kharia to restrict the sequential converb to clauses which have the same "subject" (= S/A)⁵⁰ as the main clause, as in the previous examples, which all involve a direct case subject. However, as with many other South Asian languages, a so-called "dative subject" (6.3.4) involving an experiencer marked for the oblique case may also be the controller of the sequential converb:

235. *etwa=te u=ki=yaʔ haleit yo=kon lebui laʔ=ki.*
 Etwa=OBL this=PL=GEN condition see=SEQ love EMOT=MID.PST
 'Etwa, *seeing* their condition, felt compassion.' [RD, 2:108]

If however, the A or S of the converb is not identical with the subject or experiencer (whether direct or oblique case), the sequential converb is often not used but some other construction, such as the masdar marked for the oblique case:

236. *muda dada=ḍom=ki=yaʔ aw-aw=te ber*
 but elder.brother=3POSS=PL=GEN QUAL-RDP=OBL who
beṭi=ki=te um uḍuṅ=na pal=e
 daughter=PL=OBL NEG make.flee=INF be.able=ACT.IRR
 'But as long as their elder brothers *are* present, no one will be able
 to drive the daughters away.' [MT, 1:169]

⁵⁰ Cf. footnote 1 in 6.2 for a definition of S and A.

However, at least for most speakers this is only a tendency and the use of the sequential converb is not categorically restricted to “same-subject” contexts. In the following example, it is the porcupine (*jijray*) who has stayed a few days, before the father sends his daughter off with the porcupine:

237. *thor to? aw=kon jijray=a? sori daj goth=o?*
 few day stay=SEQ porcupine=GEN with send C:TEL=ACT.PST
 ‘[Ø_i] *having stayed* a few days, [he_j] sent [her_k] off with the porcupine.’ [HJPa:40]

The use of the sequential converb in a clause whose “subject” is different from that of the main clause is, however, uncommon. In fact, for some speakers the sequential converb cannot be used in different-subject environments and the attested example (237) above was considered ungrammatical by one speaker. For other speakers, however, such examples are both acceptable and are also occasionally employed in their own texts, as in the following examples:

238. *beṭ=dom kosu buṅ jhalay=kon jhalaykon kamu=na=jo*
 son=3POSS sickness INST trouble=SEQ RDP work=INF=ADD
um ter=na la?=ki.
 NEG allow=INF IPFV=MID.PST
 ‘With her son *suffering so much* from the illness, [she] didn’t let [him] work.’ [RD, 2:104]

239. *u=gḥay gam=kan mane ho=kaṛ raj karay=na suru*
 this=way say=SEQ umh that=SG.HUM ruling do=INF begin
goth=o?
 C:TEL=ACT.PST
 ‘[They] *having spoken* thus, umh, he began to rule.’ [MS, 1:185]

240. *tama rāci eriya=te ḍel=kan piṭauriyagarh=jo*
 now Ranchi area=OBL come=SEQ Pitauriagarh=ADD
hoy=ki, ...
 become=MID.PST
 ‘Now, [they] *having arrived* in the Ranchi area, they founded Pitauriagarh (= Pitauriagarh also became), [MS, 1:213]

The converbal predicate is usually negated by *enem* 'without' (241), less often by *um* 'NEG' (242):

241. *jaisan musa in enem yo=kon saṛak tij baḍke=na*
 CREL today 1SG without see=SEQ road side hurry=INF

laʔ=ki=jn noʔ

IPFV=MID.PST=1SG Q

'Like I started running towards the street today *without looking?*'

[HJPa:95, section 4]

242. *αʔb modi iskul um col=kon a=tiʔj atiʔj peṭhiya*
 now Modi school NEG go=SEQ CREL=side REP market

kuʔdhiṅ ho=tiʔj hotiʔj co=na laʔ=ki.

fair that=side REP go=INF IPFV=MID.PST

'Now Modi, *instead of going* to school, would go wherever there was a market or fair.'

[RD, 2:26]

With respect to scope, the converbal predicate is generally within the scope of the operators marked on the finite TAM/PERSON-syntagma and shares with it the same TAM values, as the previous examples demonstrate. However, as the following example shows, this is not necessarily the case, as the predicate marked by the converbal enclitic in this example shares neither the temporal nor the aspectual values of the main predicate.

243. *raṭa jal boʔ=te col=ta ro yo=te la[ʔ] ubar*
 Rata net place=OBL go=MID.PRS and see=ACT.PRS then two

koleʔ kundū jal=te bajhe=kon saṅghra=yaʔ thoṅ
 parrot child net=OBL get.trapped=SEQ help=GEN for

īyam=taʔj=kiyar.

CRY=MID.PROG=DU

'Rata goes to the net and sees then [that] two baby parrots *have gotten trapped* in the net and *are crying* for help.'

[BB, 2:36]

The following example, an example of a so-called "tail-head linkage", also shows that the converbally marked predicate may refer to known, topical information whereas the finite TAM/PERSON-syntagma is part of the focus domain. Hence here as well, the converb is not within the scope of these operators.

244. *uɖ=na loʔɖho aɖi ho maha daru tuta=te=ga*
 drink=INF after ANAPH that big tree bottom=OBL=FOC

ɖel=ki.

come= MID. PST

‘After drinking, he *came* to the bottom of that big tree.’

ɖel=ke yo=te laʔ ho=te=ko behar=jo
 come=SEQ see=ACT.PRS then that=OBL(= ‘there’)=CNTR who=ADD

umboʔiʔj=may.

NEG.PRS.QUAL=3PL

‘*Having come*, he sees that there is no one there.’ [AK, 1:23–24]

Recall also from the discussion above that the converb may be marked for negation, independently of the finite TAM/PERSON-syntagma ((241), (242)). As obligatory sharing of such TAM and negational “operators” is often considered the defining characteristic of cosubordination as opposed to subordination, in which operator sharing is not obligatory (cf. e.g. Van Valin & LaPolla, 1997: 455), converbal clauses in Kharia are then subordinated clauses, not cosubordinated. In addition, although the majority of examples in our corpus would seem to involve core-level junctures, as both clauses share the same “subject”, examples such as (238)–(240) show that the use of the converb is not restricted to core-level junctures but is also often found in clause-level junctures.

The sequential converb in Kharia would appear to be of fairly recent origin, hence it seems reasonable to assume that until recently there was no sequential converb in Kharia and the sequentiality of related events was expressed by a TAM/PERSON-syntagma with multiple semantic sub-bases, as discussed in 6.3.2 and 6.6.1.⁵¹

6.6.2.4 Imperfective converbs (“present participles”)

There are four constructions which fulfill roughly the same functions in Kharia as imperfective converbs in other languages in that they denote that one event occurs simultaneously with the eventuality of the main predicate. In the first three constructions, we find some form of overt morphological marking following the semantic base:

⁵¹ For the expression of unrelated sequential events (‘after’, etc.), cf. 7.5.1.3.

- a form which is homophonous with the present general imperfective, middle marker =*ta*
- the focal marker =*ga*.
- the infinitival marker =*na*;

This form is then repeated. These three strategies are found with both mono- and polysyllabic contentive morphemes.

The fourth strategy is restricted to polysyllabic morphemes. Here, the lexical morpheme is repeated and no overt morphological marking is found.

Only =*ta* may unambiguously be considered an imperfective converb marker, whereas the others primarily fulfill other functions: =*ga* otherwise functions as a focus particle and is also used to express semantic depictives (7.7.4 and 7.5.1.6). Similarly, an “imperfective converb” marked by =*na* can appear in the oblique case, showing that this is in fact the infinitival marker.

The status of =*ta* as a converbal marker is much clearer: Since the present general imperfective active ending =*te* can never be used in this function and since the form marked by =*ta* cannot mark for person, this =*ta* is not the same marker as the marker of the present, general imperfective, middle, although it presumably derives from that marker.

As the following examples show, in addition to simultaneity, the “imperfective converbs” can also express an additional adverbial meanings, generally that of manner:

with =*ta*

245. *ele am=pe=te goʔ=ta goʔta han=tiʔj*
 1PL.EXCL 2=2PL=OBL carry.on.shoulders=CVB REP that=side
u=tiʔj qor=e=le.
 this=side take=ACT.IRR=1PL.EXCL
 ‘We will carry you around on our shoulders (= *carrying*, we will take).’ [AK, 2:34]

246. *u=tiʔj=ko ho=ki ho=te sima ulaʔ=ga*
 this=side=CNTR that=PL that=OBL(= ‘there’) *sima*.tree leaf=FOC
kumba bay=ta bayta aw=na maʔe=yoʔ=ki.
 hut build=CVB REP stay=INF begin=ACT.PST=PL
 ‘Here they began to stay, *building* huts of *sima* leaves.’

[MT, 1:242]

247. *dhāgar kongher daru=te luku yo=ta yota lere?=na*
 servant boy tree=OBL fruit see=CVB REP rejoice=INF
la?=ki.

IPFV=MID.PST

‘The servant boy, *seeing* the fruit on the tree, began to rejoice.’

[BB, 1:23]

with =ga

248. *muda ekle aw=ga awga muruk ansa la?=ki.*
 but alone QUAL=FOC REP very unhappiness EMOT=MID.PST
 ‘But [He (= God)] became unhappy being alone (= *being* alone, He became unhappy).’

[AK, 3:6]

249. *ro ho kuda kolon=a? daru sumbho?=te ijam=ga*
 and that millet bread=GEN tree base=OBL cry=FOC
ijamga goj jom=ta.

REP die AUTOPOES=MID.PRS

‘And *crying and crying*, she just dies [through sorrow] at the base of that millet bread tree.’

[BB, 1:90]

with =na

250. ... *lekin lam=na lamna sou²b=te ikud jughay da?*
 but search=INF REP all=OBL very much water
piyas la?=ki.

thirst EMOT=MID.PST

‘But *searching and searching*, [they] all became very thirsty.’

[AK, 1:17]

251. *ho “kayebār!” gam=na gamna=te=ga ho ghaṭo=wa?*
 that “kayebār!” say=INF REP=OBL=FOC that valley=GEN

jimi hoy guḍ=ki “kayebār ghaṭo”.

name become C:TEL=MID.PST Kayebār valley

‘By *constantly saying* “Kayebār” (= ‘pick up’), the name of that valley became “Khybar Pass (= valley)”’

[MT, 1:43]

No overt marking, only repetition

252. *raṭa=te* ḍoko ḍoko *leme²ḍ* *la?²=ki*.
 Rata=OBL sit.down REP sleep EMOT=Mid.PST
 ‘While he was seated, Rata became tired.’ [BB, 2:26]

Similarly, the repeated form of the contentive morpheme may be an “echo-word”:

253. *ho=ki* *poḍa=ḍom=ki=ya?* *halet* *onḍor=o?²=ki*, *kunḍu?*
 that=PL village=3POSS=PL=GEN condition hear=ACT.PST=PL child
hakon=ḍom=ki=te *sumtay=o?²=ki* *ro* *kuda* *kudi*
 ECHO=3POSS=PL=OBL gather=ACT.PST=PL and hurry ECHO
khaṛiya *ḍay=ki=ya?* *kunḍa²b* *kunḍa²b* *yar=o?²=ki*.
 Kharia woman=PL=GEN back REP flee=ACT.PST=PL
 ‘They heard what was happening in the villages (= the condition of their villages), gathered their children and *hurriedly* fled after the Kharia women.’ [MT, 1:218]

Alternatively, in the same construction the “echo-word” may be separated from the primary contentive morpheme by the complementizer *no*:

254. *musniṅ* *ho=ki* *jhaṛi* *kuru?* *hakon=ki=te* *somtay=o?²=ki*
 one.day that=PL all child ECHO=PL=OBL collect=ACT.PST=PL
ro *kuda* *no* *kudi* *raylogaṛh* *tay* *yar* *kan=o?²=ki*.
 and hurry Cmpl ECHO Railogarh ABL flee CONT=ACT.PST=PL
 ‘One day they gathered up all the children and fled quickly (= *hurrying*) from Railogarh.’ [MT, 1:105]

6.6.2.5 Participles

There are two unambiguously participial markers in Kharia, both of which have been borrowed from Indo-Aryan. They are only found in “prenominal” propositional attribution (7.6.2). There is also a derivational marker, *-ḍu?*, whose exact function is unclear but which can also fulfill functions similar to those of the participial markers.

- The first participial marker has been borrowed from Hindi and consists of the infinitive in *=na* followed by the form *=wala*. This form is used to denote iterativity and habituality:

255. *kamu=na=wala lebu=ki* 'working people'
 work=INF=PTCP person=PL

- The second participial marker, the suffix *-l*, attaches only to contentive morphemes borrowed from Sadri which end in *-a* or *-ay*, e.g. *tajay* 'hang (TR)' *tajay-l* 'hung', *biha* 'marry' *biha-l* 'married'. It seems to be restricted to past events only and is not acceptable to all speakers.
- The last "participial" marker, which has not been borrowed from Indo-Aryan, consists of the contentive morpheme plus the ending *-du?*, which has a number of other functions.⁵² It has no inherent temporal orientation and may refer to past, present or future events.

256. *dilli col-du? lebu* 'the man / person who went to
 Delhi go-PTCP(?) man / person Delhi'

For a full discussion of the attributive function of these "participial" forms with examples, see Section 7.6.2.

6.7 NEGATION

There are two major types of clausal or predicative negation in Kharia, depending on the status of the TAM/PERSON-syntagma, and a further major productive type used in "emphatic" negation. If the TAM/PERSON-syntagma appears in any category other than the optative or in the irrealis with an imperative interpretation, the TAM/PERSON-syntagma is negated by *um*. The imperative use of the irrealis and the optative are negated by the marker *abu*. Emphatic negation is expressed by *umbo?*. There is also a minor type of negation which is rather seldom and which has been borrowed from Indo-Aryan. These are dealt with individually in the following.⁵³

⁵² For example, as a non-productive derivational suffix of indeterminate meaning: *konsel* 'girl' vs. *konsel-du?* 'woman'. With actions its meaning is more predictable and denotes a prolonged action: *yo* 'see', *yo-du?* 'look at, stare', although it is not compatible with many contentive morphemes with this meaning. With typical property-denoting morphemes, it is highly productive and its meaning is entirely predicatable: "-ish", e.g. *rusuŋ* 'red', *rusuŋ-du?* 'reddish'. Hence, its status as a participial marker is uncertain. On the possible relation to the Juang "article" *-de* (Matson, 1964: 44), cf. fn.6, Section 3.3.

⁵³ The present section deals only with negation in primary predication. For negation of the converbs, see the section dealing with the respective converb.

6.7.1 *The general negator um*

Negation of TAM/PERSON-syntagmas other than irrealis-marked forms with an imperative interpretation or the optative is expressed through *um*, the general negative marker in Kharia.

In all persons other than the second-person singular (and occasionally there as well), when the TAM/PERSON-syntagma is negated, PERS/NUM/HON marking attaches to the negative morpheme *um*, which precedes the now partially finite TAM/PERSON-syntagma, as Table 6.49 shows, followed by a simple example. The negative morpheme *um* is not a verb or auxiliary and cannot mark for any of the TAM categories given in Table 6.49.

Table 6.49: A (simplified) schematic overview of the slots of the negated TAM/PERSON-syntagma

NEG=PERS/NUM/HON REC CAUS-LEXEME -<CAUS> V2(S)=PERF=TAM/VOICE

257. <i>ter[=e]=ijn</i>	<i>um=ijn</i>	<i>ter=e</i>
give=ACT.IRR=1SG	NEG=1SG	give=ACT.IRR
'I will give'		'I will not give'

The only (partial) exception to this is the second person, singular, where person marking may either attach to the TAM/PERSON-syntagma or to *um*, although the predicate-final position of person marking is by far the more common of the two here. Cf. (from Malhotra, 1982: 285):

258. <i>ubhroj</i>	<i>um=em</i>	<i>qe=na</i>	or	<i>ubhroj</i>	<i>um</i>
these.days	NEG=2SG	come=MID.IRR		these.days	NEG
		<i>qe=na=m.</i>			
		come=MID.IRR=2SG			
		'These days you do not come.'			

The form **um=ki* 'NEG=PL' is ungrammatical for the 3rd person, plural. Instead, only the form *umay* from **um=may* 'NEG=3PL' occurs.

259. <i>juŋ=o?=may</i>	<i>umay</i>	<i>juŋ=o?</i>	<i>kuy=o?=ki</i>
ask=ACT.PST=3PL	NEG.3PL	ask=ACT.PST	find=ACT.PST=PL
'they asked'		'they didn't ask'	'they found'
	<i>umay</i>	<i>kuy=o?</i>	
	NEG.3PL	find=ACT.PST	
		'they didn't find'	

This (highly limited) freedom of movement of PERS/NUM/HON-marking is in line with the analysis of this marker as the subject of the clause, while the “overt subject” is actually in apposition to this. See also Section 7.1.

um is also used to negate contentive morphemes in general, where it appears before the respective morpheme.

260. *dharmi ro um dharmi=ya? thom*
 righteous and NEG righteous=GEN for
 ‘for the righteous and unrighteous’

Malhotra (1982: 283, adapted here) provides a number of other similar forms, some of which are given in the following example:

261. *um lere?* ‘unhappy’ (*lere?* ‘joy; joyous; rejoice’)
um su?kho?-bo? ‘uncomfortable’ (*sukho?* ‘pleasure; happiness;
 happy’, *-bo?* ‘INTENS’)
um bes ‘bad’ (*bes* ‘good’)

Speakers felt rather uncomfortable with *um* negating units other than the finite TAM/PERSON-syntagma. For example, although they did accept the (attested) form *um dharmi* in (260), one speaker indicated that such a form is not incorrect but not used, whereas another felt *enem dharmi* ‘without righteous(ness)’ would be more appropriate. Other forms which were accepted at least by some speakers were *um suru* ‘non-beginning’, *um rusuj* ‘not red’, *um ontu?d* ‘non-end’ and, for a few, *um jhelob* ‘not long’, although the only one of these forms which met with universal approval was *um rusuj* ‘not red’, while *um jhelob* ‘not long’ met with the least approval. On the other hand, non-scalar contentive morphemes were refused outright: **um lebu* ‘non-human’, **um jinis* ‘not a thing’ or **um daru* ‘not a tree’. Clearly, the use of *um* with scalar contentive morphemes is more acceptable, although as the marginally acceptable *um jhelob* ‘not long’ shows, other factors would also seem to play a role. More research is necessary here.

The present-tense qualitative predicative (≈ “copula”, 7.3) marker has suppletive negative forms: The inherent qualitative predicative marker, *heke*, has the negative form *nalage* and the non-inherent form *ayi?j* is negated by *umbo?i?j*. Nonetheless, the following two forms are given in Malhotra (1982: 286) of the inherent qualitative predicative marker *heke* negated with *um* (although an example with the form *nalage* is also found

on the following page). As these are the only two examples in our data, it is not possible to say at the moment whether they are dialectal or idiolectal forms.

262. *am kisʔo lebu um=em heke.* 'You are not a rich man.'
 2SG rich man NEG=2SG QUAL.PRS

263. *am magra=? kanʔaybo? um heke=m.* 'You are not
 2SG Magra=GEN wife NEG QUAL.PRS=2SG Magra's wife.'

Non-present-tense forms of the qualitative predicative marker are formed with the morpheme *aw* 'live, remain, stay' These are all negated regularly with *um*.

Periphrastic TAM/PERSON-syntagmas containing an infinitive and an "auxiliary" may place the negative particle with PERS/NUM/HON marking either before or after the infinitive:

264. *ap=ʔom raʔa=te remakh=o? ro gam=o? no*
 father=3POSS Rata=OBL call=ACT.PST and say=ACT.PST Cmpl
"babu musa iʔn kinir co=na um=iʔn pal=e."
 child today 1SG forest go=INF NEG=1SG be.able=ACT.IRR
 'His father called Rata and said to him "Son, today I *will not be able*
to go to the forest." [BB, 2:21]

265. *iʔn am=aʔ sori um=iʔn eboʔ=na pal=e.*
 1SG 2SG=GEN with NEG=1SG play=INF be.able=ACT.IRR
 'I *won't be able to play* with you.' [TK, 2:21]

266. *ʔaʔhiyal=ki ho=ki=te ʔoko=na umay ter=oʔ.*
 Darhiyal=PL that=PL=OBL sit.down=INF NEG.3PL allow=ACT.PST
 'The Darhial *did not let them settle* [there].' [MT, 1:79]

267. *raʔa gam=te "iʔn u ikuʔd sundar konteʔ=ki=te*
 Rata say=ACT.PRS 1SG this very beautiful bird=PL=OBL
bajhay=kon satay=na um=iʔn lam=te."
 trap=SEQ torment=INF NEG=1SG want=ACT.PRS
 'Rata says "*I don't want to trap* these beautiful birds and *torment*
them."' [BB, 2:43]

268. *boker boksel=dom=ki buŋ teinko=jo bana*
 brother.in.law sister.in.law=3POSS=PL INST somewhat=ADD liking
um aw=na laʔ=ki.
 NEG QUAL=INF IPFV=MID.PST
 ‘She also strongly disliked her brothers and sisters-in-law (= some liking *was not*).’ [RD, 2:49]

6.7.2 Modal negation with *abu*

To negate the imperative or the optative, the form *abu* is used. The enclitic subject marker attaches to the TAM/PERSON-syntagma following *abu* or, alternatively in the second persons, *abu* itself is marked for person and number and has the forms *abu* ‘2SG’, *aʔbar* ‘2DU/2HON’ and *aʔpe* ‘2PL’. The form *abu* is also used for the 3rd persons.

269. *muda u gharana=? jimi uʔbne toʔ purluʔd purluʔd*
 but this family=GEN name so.much day spotless REP
ayiʔj jeʔ abu koɖil guŋuʔ.
 QUAL.PRS 3SG.NHUM NEG.MOD stain OPT
 ‘But this family’s name has been spotless for such a long time, [he] *should not stain* [it].’ [Kerkeṭtā, 1990: 12]
270. *muda soloʔ=te=ko lemedʔ=ga abu de=na.*
 but dog=OBL=CNTR sleep=FOC NEG.MOD come=MID.IRR
 ‘But the dog just couldn’t get any sleep (= but to the dog, sleep *should not come*).’ [BB, 3:36]

Not all forms which appear in the irrealis are negated with *abu* but only those with an imperative sense:

271. *brahman, musa tay am doli=te um*
 Brahman today ABL 2SG palanquin=OBL NEG
gog=e=m,
 carry=ACT.IRR=2SG
 ‘Brahman, as of today *you will not carry* the palanquin,’ [AK, 2:31]

272. *kongher=te lebuy la?=ta ro gam=te* "abu
 boy=OBL pity EMOT=MID.PRS and say=ACT.PRS NEG.MOD.2SG
ijnam=na, yaya "
 cry=MID.IRR grandmother
 'The boy feels pity and says "Do not cry, Grandmother."'
 [BB, 1:58]

273. *kay=e=bar! kimin kũru?, a?=bar=ga ro?*
 lift=ACT.IRR=2HON daughter.in.law child NEG.MOD=2HON=FOC drop
melay=e!
 leave=ACT.IRR
 'Pick [them] up (HON)! Daughter-in-law, *don't drop [and] leave*
[any]!' [MT, 1:42]

As noted, *abu* in the second persons does not always mark for person or number. Instead, this marking is often found on the TAM/PERSON-syntagma, as in the following example. This pattern is in fact quite common.

274. *i lam=te=m? la?, mane, ijn ter[=e]=ijn*
 what want=ACT.PRS=2SG then umh 1sg give=ACT.IRR=1sg
lekin ijn=te jan abu tar=e=m.
 but 1sg=OBL life NEG.MOD kill=ACT.IRR=2sg
 'What do you want? Then, umh, I will give [it to you], but *don't*
kill me.' [MS, 1:64]

6.7.3 Emphatic negation with *umbo?*

The form *umbo?* 'no; not' can be used to express emphatic negation. As the following example shows, it may appear at the end of the clause directly after the finite TAM/PERSON-syntagma.

275. *jijray diyo=ga laro? daru=te yo=na co=na*
 porcupine daily=FOC date.palm tree=OBL see=INF go=INF
la?=ki, muda goj=ta umbo?
 IPFV=MID.PST but die=MID.PRS no
 'The porcupine went to see the date-palm tree daily, but it *just*
wouldn't die.' [HJpa:39]

It is also often found in place of the negated qualitative predicative markers (\approx copulas):

276. *joʔ=na no uʔdʔ=na=yaʔ tʰekan umboʔ.*
 eat=INF or drink=INF=GEN fixed.place no
 ‘There *was no fixed place* for eating or drinking.’ [RD, 2:53]

Negation with *umboʔ* is also found with the masdar, especially in contexts where most of the information in the clause is already pragmatically active. Malhotra (1982: 288) writes that this use is restricted to first-person forms only and with a “capabilitative passive” meaning (= “could not”), for which she gives three examples:

277. *umboʔ yo-yo*⁵⁴ ‘I could not see it.’
 NEG see-RDP

278. *umboʔ col-col* ‘I could not go’
 NEG go-RDP

279. *umboʔ joʔ-joʔ* ‘I could not eat’
 NEG eat-RDP

However, as the example she cites on the previous page (from Pinnow, 1965a: 38), given below, shows, this is clearly not the case as it is compatible with any persons. Also, there is no “capabilitative” meaning here.

280. *kandʔayboʔ koŋ-kon umboʔ, ho=kaʔ goʔ=kon*
 old.woman know-RDP NEG that=SG.HUM carry=SEQ

dʔe=na laʔ=ki.
 come=INF IPFV=MID.PST
 ‘The old woman *did not know* [this], she came carrying [him].’
 [HPa:38]

Nevertheless, Malhotra’s insight—at least with respect to person—is valid as a tendency, and speakers often indicated in interviews that only

⁵⁴ Mistakenly given as *umbo yoʔ-yoʔ*. Although the form *umbo* is often encountered, as the glottal stop is often omitted in normal speech, the form *yoʔ-yoʔ* is clearly a misprint.

'I' and 'we' (DU/PL, INCL/EXCL) could be interpreted as the person who the negated clause refers to, although counter-examples are quite easy to find. This would seem to result from the fact that this type of negation is especially common (although not obligatory) in answers to questions, and with certain contentive morphemes—such as *koŋ* 'know'—the answer *koŋ-koŋ umbo?* 'don't know' will generally be interpreted as referring to the speaker (similar to the English translation just given). However, this is only tendentially so, as (280) shows.

6.7.4 Negation with *na*

Very rarely, the Indo-Aryan loan word *na* 'NEG' is found, usually either in negative commands, with the infinitive (281), or in *neither...nor* expressions with a morphologically finite TAM/PERSON-syntagma (282).

281. *la?* *am* *na* *cori* *karay=na*, *na* *jahãy=a?* *jan*
 then 2SG NEG theft do=INF NEG INDEF.HUM=GEN life
tar=na, *na* *jahãy=a?* *jahã* *karay=na*.
 kill=INF NEG INDEF.HUM=GEN INDEF.NHUM do=INF
 'So *do not steal, do not kill anyone*, do not harm anyone (= *do not do anyone's anything*).' [MS, 1:309]

282. *modi* *na* *silo?=na* *la?=ki* *no* *oḍo?=ga* *hoḍom*
 Modi neither plow=INF IPFV=MID.PST or more=FOC other
kamu *karay=na* *la?=ki*.
 work do=INF IPFV=MID.PST
 'Modi *neither plowed* nor did any other work.' [RD, 2:39]

6.8 PERIPHRASTIC TAM/PERSON-SYNTAGMAS ("VERBS")

There are a number of complex TAM/PERSON-syntagma types which consist of an infinitival element containing lexical information and a finite functional head, which will be referred to here for convenience as an "auxiliary" These cover a variety of functions, including aspect in its most general sense ("phase verbs" and general imperfectivity) as well as modals.

6.8.1 “Aspectual” auxiliaries

6.8.1.1 *la?* ‘imperfectivity’

This is by far the most common auxiliary. It is usually found in the past tense and is also compatible with the irrealis, the present general imperfective or with the perfect, but not with the present progressive. It denotes imperfectivity in general, including iterativity, habituality and non-present progressivity. It is always marked for the middle voice in those categories in which an active / middle opposition is found (i.e., simple past, present, irrealis).

283. ... *khariya lebu=ki pujapaṭh karay=na la?=ki=may, ...*
 Kharia man=PL sacrifice do=INF IPFV=MID.PST=3PL
 the Kharia men *used to perform* sacrifices. [AK, 2:6]

284. *um la?=ko um uḍ=na la?=ta hā?niṅ.*
 NEG then=CNTR (= ‘but’) NEG drink=INF IPFV=MID.PRS right?
 ‘But otherwise *he doesn’t drink*, right?’ [Kerkeṭ[ā, 1990: 25]

285. *aḍi ukho?=ghay socay=na la? goʔd=si? bhere=ga.*
 ANAPH this=way think=INF IPFV C:TEL=PERF time=FOC
 ‘When he *has been thinking* thus.
 [Kullū, 1992: 2 = Matthew, 1:20]⁵⁵

la? seems to derive from *lag-* in Indo-Aryan, where it means ‘attach’ and is found, among other uses, with the meaning ‘begin’ when it follows an infinitive. And in fact, *la?*, although it generally denotes imperfectivity in Kharia, is occasionally found with an inceptive meaning. This double interpretation is common to many languages of Jharkhand and Bengal (cf. Peterson, forthcoming, a). Although the data for this interpretation are still too scant, it would seem that this meaning is especially common in narratives when a new episode is being presented. Cf. the following two examples.

In the first example, we have the following situation: A young boy, Rata, who is trapping birds in the forest for his father, who is sick and cannot work, falls asleep and dreams that he has been captured by birds

⁵⁵ Translation in the Saint Joseph Edition of *The New American Bible*: “Such was his intention . . .” (Catholic Biblical Association of America, 1970: The New Testament, 6).

and is being carried up to the sky to be let drop as a way of stopping his family from trapping birds. The story reads as follows. Only the relevant passage is given here in the original, the preceding and following text is given only in the English translation:

[BB, 2:28–35]

28. While he was asleep, he had a dream: The birds were carrying him towards the sky.
29. At that very moment a crow came and said to the birds: “Why are you bringing this sinful man’s child to Heaven?”
30. They replied: “We will drop him from the sky and warn his father, Darhi.
31. Because Darhi will destroy our lineage by trapping us.
32. Therefore we will drop him from above and finish off Darhi’s descendants.”
- 34.

286. *hobne=te=ga ubar kole? kunḍu? jal=te bajhe=kon*
 that.much=OBL=FOC two parrot child net=OBL get.trapped=SEQ

“*ṭāy-ṭāy*” toro³d=na laʔ=ki=kiyar.

“*tay, tay!*” cry=INF IPFV=MD.PST=DU

‘Meanwhile (= in that much), two baby parrots got caught in the net and *began crying* “Tay! Tay!”’

34. The two parrots’ mother and father keep jumping here and there to free them from the net and all the birds are crying “Tay! Tay!”’
35. Upon hearing their cries, Rata woke up.

Rata then frees the birds and decides to take up a different profession. Line 34 (= (286)) is thus a turning point in the story, and the “imperfective” here to denote ‘begin’ seems to serve to introduce the background information for the next scene, hence although this a new action in the story, i.e., although something begins here, it serves as background information for what is to come.

The following presents one further example. Again, only the relevant passage is given in the original, all else is given only in the English translation. In this story, a dog is searching for an animal to be his friend, and goes from one animal to another, but is unhappy with all he has met so

far. He has just befriended a jackal, but now considers him a coward and decides to find another friend.

[BB, 3: 44–56]

44. Having thought this the dog says to himself: “I will leave the jackal now and make the tiger my friend.”
45. And he went to look for the tiger. Searching and searching he found the tiger and says: “How are you brother, OK?”
46. The tiger says: “Very good.” And asks: “And you?”
47. The the dog says: “Without a friend I am not well.
48. I want to make you my friend, so what do you say?”
49. The tiger says: “It will be good.” And they began to live together.
50. At night they both lay down at the base of a tree.
- 51.

287. *kiŋo? lemed=sikh=o?. acka=ga saŋa onɔor=kon solo?*
 tiger sleep=PERF=ACT.PST sudden=FOC sound hear=SEQ dog

bhabru=na la?=ki.

bark=INF IPFV=MID.PST

‘The tiger fell asleep. Suddenly, having heard a noise, the dog *began to bark.*’ [BB, 3:51]

52. Having heard his cries, the tiger awoke and swore at the dog: “If you aren’t tired, then why do you annoy others?”
53. Hearing your barking, a man will come and kill us both.
54. Lie down quietly.” The dog did not like the tiger’s words.
55. And he says to himself: “The tiger is a coward, he fears the man.
56. Now I will leave the tiger and make the man my friend.”

The dog now sets out to befriend the man. Here again, the imperfective is used to signal an action which begins, however one which also serves as the background for what is to come. Further research is necessary to determine if this adequately accounts for this use of the “imperfective” marker *la?* in general.

la? perhaps originally filled a function similar to the “indefinite past” in *-jan* in Mundari, which signifies “that the Subject or Agent *went* to do a certain work or *began* some action, and that he has not yet completed or discontinued it: ” (Hoffmann, 1905 [2001]: 183, emphasis in origi-

nal). This might explain how a morpheme which originally meant 'begin' came to function primarily as a marker of the non-present imperfective. More research is necessary on this topic. Cf. also *māṛe* in the following section.

The past imperfective, which is otherwise marked by the past middle marker =*ki*, can also be marked for the "Past II" (6.4.1.1).

288. *da?* *ṣṭesan* *pase=ga* *moṅ* *daru* *sumbo?=te* *ho*
 then train.station near=FOC one tree base=OBL that
carwaha *ḍoko=kon* *mukum=na* *la?=kho?*
 shepherd sit.down=SEQ doze.off=INF IPFV=PST.II
 'Then near the train station, at the base of a tree, that shepherd sat down and *began to doze off*.' [RD, 1:8]

289. *ho=kaṛ* *ḍisa?* *ḍisa?* *kho?tay=jo* *merom* *gupa=na*
 that=SG.HUM far REP up.to=ADD goat guard=INF
la?=kho?
 IPFV=PST.II
 'He *used to tend* the goats even going (= also up to) very far.' [RD, 1:4]

6.8.1.2 *māṛe*/*maṛe*,⁵⁶ *suru* 'begin'

Both *māṛe* and *suru* combine with an infinitival form to denote the inception of an action or event. Unlike *la?*, described in 6.8.1.1, these morphemes have only this one meaning.

290. *je?* *botoṅ* *buṅ* *thartharay=na* *māṛe=yo?=ki*
 so fear INST tremble=INF begin=ACT.PST=PL
 'So [they] *began to tremble with fear*.' [MT, 1:32]
291. *hin* *bhere* *ho* *tay=jo* *yar=na*
 that time that ABL(= 'from there')=ADD flee=INF
maṛe=yo?=ki *la?* *purkha=ki* *maha* *dano=te*
 begin=ACT.PST=PL then ancestor=PL big demon=OBL

⁵⁶ Similar to the proposed etymology of *la?* from *lag-* 'begin', *māṛe* 'begin' as a contentive morpheme means 'remain, stay'. Again, a category such as the "indefinite past" in Mundari seems likely for an earlier period of Kharia, i.e., 'begin and remain doing'

tar=oʔ=ki.

kill=ACT.PST=PL

‘At that time as well they *began to flee* from there, then the ancestors killed a great demon.’ [MT, 1:71]

6.8.1.3 *melay* ‘stop’

melay in combination with an infinitive denotes that an event comes to an end.

292. *hin=aʔ ghaʔd=ga ho=ki=te oʔb-koda=na melay*
 that=GEN for=FOC that=PL=OBL CAUS-paint.oneself=INF leave

goʔh=oʔ=ki.

C:TEL=ACT.PST=PL

‘Therefore, [they] *stopped tattooing (=painting)* them.’

[MT, 1:173]

293. *soloʔ bhabru=na melay=oʔ.*

dog bark=INF stop=ACT.PST

‘The dog *stopped barking.*’

[BB, 3:24]

6.8.1.4 *col* ‘go’

col, when used in combination with an infinitive, denotes that an event is about to take place, as a kind of immediate future marker.

294. *tama iŋ aniŋi=aʔ gotar=aʔ kahni batay=na*
 now 1SG 1PL.INCL=GEN clan=GEN story tell=INF

col=taʔjɔ=iŋ.

go=MID.PROG=1SG

‘Now I *am going to tell* the story of our clan.’

[AK, 1:1]

6.8.2 *Modality*

6.8.2.1 *lam* ‘look for; want’

As an independent morpheme, *lam* has the meaning ‘look for’ As an auxiliary, it has two uses:

- With animate subjects, *lam* expresses the volition of the subject. This use was acceptable to all speakers and is the most common “auxiliary” function of *lam*.

295. *ij̄n u iku²d̄ sundar konthe²d̄=ki=te* bajhay=kon
 1SG this very beautiful bird=PL=OBL trap=SEQ
 satay=na um=ij̄n lam=te.
 torment=INF NEG=1SG want=ACT.PRS
 'I don't want to trap these beautiful birds and torment them.'
 [BB, 2:43]

– For many speakers, *lam* may also be used with inanimate subjects and then expresses something akin to 'be imminent'. Although this use is very common for some speakers, it is unacceptable to others, who said this would mean that the inanimate subject would have to want to perform the action.

296. *biru raj̄j̄ anij̄=a? purkha=ki=ya? d̄oko ro*
 Biru kingdom 1PL.INCL=GEN ancestor=PL=GEN settle and
d̄ho?-d̄ho? [h̄ãro hoq̄om=a? ti²j̄=te co=na lam=tej̄.
 take-RDP place other=GEN side=OBL go=INF want=ACT.PROG
 'Biru Kingdom, the place where our ancestors settled and which they took, is about to go to the side of another [i.e., be conquered by outsiders].'
 [Kerkeṭṭā, 1990: 6]

6.8.2.2 *ter* 'give; allow'

ter in conjunction with an infinitive has a permissive meaning:

297. *um̄bo?, j̄iyom aw=na bheir=ko ho=ghay=na um=nīj̄*
 NEG life QUAL=INF up.to=CNTR that=way=INF NEG=1PL.INCL
 ter=e.
 allow=ACT.IRR
 'No! As long as there is life [in us], we will not allow that to happen (= we will not allow [it] to become that way).'
 [Kerkeṭṭā, 1990: 7]

298. *beṭ=d̄om kosu buj̄ j̄halay=kon j̄halaykon*
 son=3POSS sickness INST be.troubled=SEQ REP
 kamu=na=jo um ter=na la?=ki.
 work=INF=ADD NEG allow=INF IPFV=MID.PST
 'With her son suffering so much from the illness, [his mother] didn't even let him work.'
 [RD, 2:104]

6.8.2.3 *pal* ‘can, be able’

With an infinitive, *pal* denotes ability. This is in contradistinction to its use as a “v2”, discussed in Section 6.5.12.4, where it directly follows the semantic base of the TAM/PERSON-syntagma and has the meaning ‘finish’

299. *paʔtopur=te=ga ho=ki ikud=ga kisro=sikh=oʔ=ki*
 Patna=OBL=FOC that=PL much=FOC wealthy=PERF=ACT.PST=PL

hin=aʔ caʔdom ho=ki rohtaspur hinte doko dhab=na
 that=GEN for that=PL Rohitasgarh LOC sit.DOWN SUD=INF

pal=oʔ=ki

be.able=ACT.PST=PL

‘In Patna they had become very wealthy, therefore they *were able to settle down quickly* in Rohitasgarh.’ [MT, 1:188]

300. *muda raksin se dinu i=jo karay=na um*
 but witch that day what=ADD do=INF NEG

pal=oʔ ro oʔ col kan=ki.

be.able=ACT.PST and house go CONT=MID.PST

‘But the witch *could do nothing* on that day and went home.’

[BB, 1:50]

6.8.2.4 *The qualitative predicative marker: Obligation and necessity*

The qualitative predicative marker (cf. 7.3) expresses obligation or necessity when used in conjunction with an infinitive. The person who is under obligation appears in the oblique case and the infinitive is the “overt subject” The qualitative predicative marker is always unmarked for person, which is formally identical to the 3rd person (singular).

301. ... *ijn=te ranga bhere=yaʔ ghaʔd terom thuray=na*
 1SG=OBL cold time=GEN for honey gather=INF

ayiʔj.

QUAL.PRS

‘... *I have to gather* honey for the cold season.’ [TK, 2:20]

Although the stative present-tense qualitative predicative markers are most commonly found in this function, any of these forms can be used in this construction, including *hoy* ‘become’:

302. *muda* am=te *ij=a?* o? co=na hoy=na.
 but 2SG=OBL 1SG=GEN house go=INF become=MID.IRR
 'But you will have to go to my house [to live].' [BB, 3:62]

6.9 INFERENCE

The primary means of expressing inference is through the sentence-final marker *hoy*, homophonous with *hoy* 'become', from which it undoubtedly derives.⁵⁷ *hoy* appears to be compatible with all TAM categories. In the corpus, *hoy* always appears sentence-finally, directly following the TAM/PERSON-syntagma. As recorded data for *hoy* are still too scant, it is not yet possible to determine whether it is a phonological word or an enclitic.

303. *ida?=ki* *del=ki* *hoy*. 'He came yesterday or so, I
 yesterday=PL come=MID.PST INFER guess.'
304. *ro* *konoḍ=te* *tay* *gam=te* *no* "jarul=ga *kiro?*
 and think=ACT.PRS then say=ACT.PRS Cmpl certain=FOC tiger
tuyu=ya? *tay* *jughay=ga* *ḍiḥgar* *ro* *sawāgar* *ayi?*
 jackal=GEN ABL much=FOC brave and strong QUAL.PRS
hoy."
 INFER
 'And he thinks and then says [to himself] "The tiger is certainly
 much braver and stronger than the jackal."' [BB, 3:43]

(seeing that the ground is wet:)

305. *ḍa?* *del=ki* *hoy*. 'It must have rained (= water must
 water come=MID.PST INFER have come).'

⁵⁷ *hoy* derives from Sadri *hoi* 's/he / it will be / become' (Nowrangi, 1956: 67). This development is also found in other languages, cf. Peterson (2000), where a similar development is argued for the quotative and inferential markers in Nepali from a copula.

306. *ad̥i* *tama* *ho?* *eŋ=ki* *hoy.* ‘He must have returned
 ANAPH NOW house return=MID.PST INFER home by now.’

hoy denotes both what are referred to in Peterson (2000) as “inference through reasoning”, in which no direct evidence is involved (e.g., (304)) as well as “inference through direct results”, as in (305), which could be uttered, e.g., after going outside in the morning, seeing that the ground is wet, and inferring that it must have rained during the night. It is not used for either mirativity (cf. Delancey, 1997) nor in quotatives, which are also often marked by the same marker as inferentials (cf. DeLancey; 1997; Lazard, 1999; Michailovsky, 1996; Peterson, 2000).

SYNTAX

7.1 GRAMMATICAL RELATIONS

In this study, the methods of the structural-typological tradition are employed when discussing grammatical relations. It is customary in this approach to examine a number of criteria, such as those discussed in Keenan's (1976) seminal study on subjects from a cross-linguistic perspective, although opinions can vary greatly on the "correct" methodology, as even a brief overview of the literature shows.¹

This section concentrates on morphological and syntactic aspects of the notion of subject. Following Sasse (1978) it is assumed that a subject must be unambiguously encoded in some manner, whether through morphology (e.g., case marking, predicative morphology), word order, or through a combination of these.² This is to be distinguished from what is often termed the "syntactic subject" Here the notion of "pivot" is used to discuss these syntactic operations individually. In order for a language to have a syntactic subject in this sense, one will need to show that there is a "privileged NP" (Foley & Van Valin, 1985) or, in our terminology, a privileged type of Case-syntagma, which is the pivot for various operations. This is necessary since even within one and the same language, different grammatical operations such as reflexivization, coreferential deletion, etc., may function along different pivots, and some of these processes may not be subject to any pivot at all.

A further complication is that we must also show that a particular Case-syntagma is indeed an argument in the first place and not merely

¹ Cf. e.g. Peterson, 1998: 67ff. for a brief overview of the definition of "subject" in Keenan (1976), Dixon (1994), Foley & Van Valin (1985) and Sasse (1978).

² Of course, there is often the further complication that even within a single language, nominal and verbal morphology may point in different directions, as in Nepali (Indo-Aryan, cf. e.g., Peterson, 2002b) or Warlpiri (Australian, Simpson, 1991) and elsewhere (cf. Palmer, 1994: 56–7). Cases such as these will have to be dealt with on an individual basis and in these cases it may not be appropriate to speak of a morphological subject at all. This will not play a role in the following discussion, however, as Kharia does not display such split morphology.

an adjunct (in its broadest sense, including apposition). In fact, this is probably the thorniest issue in Kharia and is hence the topic we begin with in the following section.

7.1.1 *Case-syntagmas and their status within the clause*

We begin our discussion here with the status of the subject, the only argument marked on the TAM/PERSON-syntagma. This marker is enclitic, among other reasons since it appears at the end of the non-negated TAM/PERSON-syntagma and usually (although not always) follows the negative marker *um* in negation (for further details, see 6.7.1). This suggests that this enclitic is in fact the subject and not an agreement marker.

1. *koŋ=te=le* ‘we know’ *um=le* *koŋ=te* ‘we don’t know’
 know=ACT.PRS=1PL.EXCL NEG=1PL.EXCL know=ACT.PRS
2. *koŋ=te=m* ‘you know’ *um=em* *koŋ=te* / *um*
 know=ACT.PRS=2SG NEG=2SG know=ACT.PRS NEG
- koŋ=te=m* ‘you don’t know’
 know=ACT.PRS=2SG

The issue is complicated somewhat, however, by the fact that the overt “subject”, in this case a proform, is compatible with the presence of this explicit subject marker:

3. *ele* *koŋ=te=le* ‘we know’
 1PL.EXCL know=ACT.PRS=1PL.EXCL
4. *am* *koŋ=te=m* ‘you know’
 2SG know=ACT.PRS=2SG

Despite the fact that the presence of the subject enclitic is compatible with the presence of an overt Case-syntagma referring to the same entity, we must nevertheless consider the subject enclitic to be the subject of the clause and not an agreement marker, since the enclitic subject is obligatory (unlike the Case-syntagma) and since as an enclitic it is also a syntactic atom which occupies its own node in the syntactic structure: If we consider the overt Case-syntagma to be the subject, the result would be a clause with two subjects. It is therefore better to consider the subject

marking on the TAM/PERSON-syntagma to be the subject itself, and the apparent “overt subject” to be in apposition to this.

As we shall see in 7.7, Case-syntagmas generally only appear either when they are focussed or when there is the possibility of ambiguity. In view of this, we analyze the “overt subject” as a bound external or “overlay” function (cf. e.g., Andrews, 1985: 82–89; Foley & Van Valin, 1985: 300–303; Falk, 2001: 59–60). This unit then refers to the subject of the clause, marked on the TAM/PERSON-syntagma, and appears in the direct case (i.e., no overt case marking). In this respect, it resembles topicalization in languages such as English.

There are also other arguments in favor of this analysis. For example, this analysis enables us to account in a systematic way for the many cases in which the formal categories of person and number are not the same for the overt Case-syntagma as for the enclitic subject marker on the TAM/PERSON-syntagma. As a Case-syntagma is merely in apposition to the subject of the clause, there is no general requirement that the two must have the exact same features, e.g., with respect to person, as long as it may be construed as referring to the same entity as the enclitic subject. Cf. the following example.

5. *botoŋ=ta=pe ho=ki lutui su=kon pe? cakhna? iʔjthaŋ*
 fear=MID.PRS=2PL that=PL clothes put.ON=SEQ rice curry cow.dung
kinbhar=na=pe, in=ko laŋe=na=in.
 courtyard=MID.IRR=2PL 1SG=CNTR fight=MID.IRR=1SG
 ‘Those of you who are afraid, you put on your [house] clothes and see to your house work like cooking and cleaning the courtyard with cowdung, but I will fight.’ [Kerkeṭṭā, 1990: 7]

Consider the apparent contradiction between the “overt subject” and the “agreement” marker on the TAM/PERSON-syntagma. The subject would appear to be *hoki* ‘they’, but the TAM/PERSON-syntagma is clearly marked as having a second-person, plural subject, =*pe* ‘you (PL)’. This is further complicated by the fact that *hoki* is the head of a “relative clause” which consists of the TAM/PERSON-syntagma *botoŋ=ta=pe* ‘you (PL) are afraid’.

As this utterance is directed toward a large group of people, the subject is the second person, plural, although only a part of this group is meant, i.e., ‘(those of) you’, not ‘(all of) you’. In cases such as these, Kharia regularly chooses the second person over the third person in terms of subject marking, as the group whose members are intended is being

directly addressed. This also entails “overriding” what would seem to be the explicitly mentioned grammatical subject, *hoki* ‘they’. If however we view this “overt subject” as being in apposition to the subject enclitic on the TAM/PERSON-syntagma, further specifying this subject in a part-whole relation, the apparent mismatch is resolved.

This last example is also not a “mistake” in any sense, as this and similar examples were discussed with a number of native speakers and all found them to be perfectly correct and natural. The following presents a few other, similar examples.

6. *korob=si?na=pe. ber=jo i=jo a?pe gam=e.*
 silent=PERF=MID.IRR=2PL who=ADD what=ADD NEG.MOD=2PL say=ACT.IRR
 ‘Be quiet! *Don’t any of you say anything.*’ [Kerke[[ā, 1990: 2]
7. *behar bujhay=e=pe?* ‘Which of you will explain?’
 who explain=ACT.IRR=2PL
8. *ghaq ađi je bhere eñ=ki r[o] aba=dōm=te sou?b*
 therefore ANAPH CREL time return=MID.PST and father=3POSS=OBL all
bhai=kiyar=a? sori? potom=te kholay kholay ob-yo=na
 brother=HON=GEN together bundle=OBL open REP CAUS-see=INF
la?=ki=may, se bhere ađi=ya? potom=te soreñ kui=ki.
 IPFV=MID.PST=3PL that time ANAPH=GEN bundle=OBL stone find=MID.PST
 ‘Therefore, when **he** returned, and **along with all his brothers** [*they*]
 opened the bundles and *were showing* them to their father, at that
 time he found a stone in his own bundle.’ [AK, 1:72f.]
9. *moñ khōṛha kayebar ghaṭo tay eqo?=ga to?bluñ del=ki=may.*
 one section Khyber valley ABL more=FOC high come=MID.PST=3PL
 ‘**One section [of the group] came** (PL!) up even higher than the
 Khyber Pass.’ [MT, 1:48]
10. *babu, kimin=ki jahā no?=na cij dho?=si?=pe*
 child daughter.in.law=PL INDEF.NHUM eat=INF thing grab=PERF=2PL
gur=na, . . .
 fall=MID.IRR
 ‘Child, some of the food (= things to eat) *which [you] daughters-in-*
law have taken will fall, [MS, 1:44]

In the following example, the daughter is being asked if she will meet with her fiancé, Dele, who is not present in this discussion. Speakers stated that the subject enclitic =*bar* here is not the second person, honorific but rather the second person dual, as it refers to both members of the couple.

11. *ḍele buṅ kol kui=na=bar? bariya=ga gorjhuṅ*
 Dele COM REC find=MID.IRR=2DU both=FOC path

yoku²j=na=le.

wait.for=MID.IRR=1PL.EXCL

‘Will [**you (sg!)**] *meet up* (= *DU*) with Dele? We will be expecting [you] both.’ [Kerkeṭṭā, 1990: 16]

It could of course be argued that examples such as these are found in any language, such as English (*A number of problems arise / arises*) or the so-called *constructio ad sensum* in classical languages where semantic considerations would seem to override morphology and are usually considered “performance errors” in “lively conversation” The difference to these is that in Kharia this construction is not limited to individual contentive morphemes or constructions and may not be considered performance errors. Rather, it is a fully grammatical pattern in Kharia and in interviews speakers saw nothing unusual in this type of marking. Finally, the fact that a number of these examples are from Kerkeṭṭā (1990), a published work, clearly shows that they are not “performance errors”

Furthermore, what would appear to be the “overt subject” need not even have the form of a Case-syntagma but may also be a TAM/PERSON-syntagma. Thus, although the two have a very close relation to the functional categories of reference and predication, respectively, they cannot be equated with these:

12. *kunḍab aw=ki tomliṅ khariya gam ḍom=na la?=ki=may*
 behind QUAL=MID.PST milk Kharia say PASS=INF IPFV=MID.PST=3PL

ina no u=ki tomliṅ u²ḍ=ga ḍel=ki=may.

because this=PL milk drink=FOC come=MID.PST=3PL

‘[Those who] were in the rear were called “Milk Kharia” because they came drinking milk.’ [MT, 1:180]

It would of course be possible to assume a “zero proform” *hoki* ‘they’ here, however the present analysis, in which this is only in apposition to

the subject =*may* ‘3PL’ on the TAM/PERSON-syntagma, would seem preferable in light of the data given above.

We therefore conclude that the enclitic subject marking on the TAM/PERSON-syntagma is in fact the subject of the clause and that the overt Case-syntagma, if present, is merely in apposition to this. This is also in line with the enclitic status of the subject marker as a syntactic atom.

The question now remains with respect to the other overtly mentioned Case-syntagmas in the clause, and the status of these elements is less straight-forward than that of the subject. Note that none of these are marked on the TAM/PERSON-syntagma, however these “arguments” are also never obligatory. Consider the following three examples:

13. *ho=kay=aʔ thoŋ caha bay kay=oʔj.*
 that=SG.HUM=GEN for tea make BEN=ACT.PST.1SG
 ‘I made tea for him/her.’

14. *caha bay kay=oʔj.*
 tea make BEN=ACT.PST.1SG
 ‘I made tea for [him, her, us, you, someone].’

15. *ho=kay=aʔ thoŋ caha bay=oʔj.*
 that=SG.HUM=GEN for tea make=ACT.PST.1SG
 ‘I made tea for him/her.’

In (13) the TAM/PERSON-syntagma is overtly marked as having a benefactive participant and at first sight it might seem reasonable to term this unit an argument which has been incorporated into the predicate structure through the benefactive v2 *kay* as a kind of “applicative”. However, (14) shows that the presence of this “argument” is not obligatory, as it may be “omitted” if its identity is clear from context or is considered unimportant. In view of this, we would probably have to speak here of “pro-drop” to account for this data.

(15) however calls this analysis into question: Here, the same “argument” is explicitly mentioned and has the same form as in (13) but its status as an argument is called into question by the fact that its presence is not overtly expressed in the TAM/PERSON-syntagma. In (15) we would seem to have an adjunct, despite the fact that it receives the same marking as the “argument” in (13).

In fact, no overt Case-syntagma in the clause is ever obligatory, as the following examples from interviews show. Note that these are situations

in which no information may be presumed, and yet no “arguments” are obligatory.

(Upon entering a room in a state of disorder:)

16. Q: *ute i hoy=ki?* A: *surum=oʔ=ki.*
 here what become=MID.PST steal=ACT.PST=PL
 ‘Q: What happened here? A: They stole.’

(Another situation, also no presupposed information, other than the identity of the subject:)

17. Q: *laʔ i karay=oʔ ?* A: *māṛay=oʔ.*
 then what do=ACT.PST put=ACT.PST
 ‘Q: Then what did s/he do? A: S/he put.’

Examples such as these were posed to six different speakers in interviews in which none of the situations or participants in the examples, other than the subject in (17), could be considered presupposed information and therefore omitted since their identities were known. Although this was of course an interview situation and this may affect the results somewhat, it is significant that not one of the six speakers had the slightest objection to these two examples or other similar ones. All simply thought up situations in which such sentences could be uttered. Upon being asked how such utterances could be interpreted in the total lack of any knowledge as to the identities of the participants, one speaker simply noted “If you don’t know what they’re talking about, you just ask.”

Data such as these strongly suggest that all clauses in Kharia are syntactically intransitive:³ The only obligatory argument in Kharia is the enclitic subject marker on the TAM/PERSON-syntagma. All overt Case-syntagmas are bound external functions, similar to the discussion of the “overt subject” above, with the exception that units which are not in apposition to the (enclitic) subject are case marked for their respective functions in the clause. Thus, whereas the “overt subject” in Kharia closely resembles topicalization in English and obligatorily refers to the subject marker

³ As Hengeveld & Rijkhoff (2005:426ff.) note, languages which do not possess nouns and verbs as lexical categories also at least tend to have only intransitive contentive morphemes. This fits in well with the Kharia data, cf. Chapter 4.

on the TAM/PERSON-syntagma, other Case-syntagmas are more similar to left-dislocation in that they are not coreferential with any element in the core.⁴

In this view, the core of the clause consists of the TAM/PERSON-syntagma and its subject marking—all Case-syntagmas contained within the clause are core-external. With that, the sentence in Kharia consists of two principal units: The first (optional) unit is a kind of “prefield” which serves as an orientation for the hearer and provides referential information as to the participants and their roles in the clause, if this is deemed necessary.⁵ The second part consists of the TAM/PERSON-syntagma itself, the core. It is here that an event or state is narrated and the identity of the subject is given and which specifies the number of participants in the scenario: In addition to the subject marker, the TAM/PERSON-syntagma also unambiguously indicates whether or not an object is involved: Contentive morphemes are either unambiguously intransitive or transitive, semantically speaking, or if they may be both transitive and intransitive, the use of the active will signal the presence of an object at the semantic level whereas the use of the middle voice signals the presence of only the subject, at least outside the generic function of the middle voice. Other entities, such as a benefactive, may also be signalled on the TAM/PERSON-syntagma. As their presence at the semantic level (although not their identity) is signalled on the TAM/PERSON-syntagma, this unit refers to this semantic argument but is itself not an argument, similar to the “overt subject” and “overt object” Finally, there is an optional postfield following the TAM/PERSON-syntagma, which is much less common than the prefield (cf. the discussion in 7.7).⁶

It must be stressed here that the active and middle signal SEMANTIC transitivity, not morphosyntactic transitivity. As we saw above, NO overt

⁴ I must stress that the model I am developing here for Kharia RESEMBLES topicalization and left-dislocation in English. This is not to say that it is identical to these two operations in English or elsewhere, neither with regard to their functions nor structurally. For example, with regard to structure, left-dislocation in English arguably involves the movement of a constituent to the beginning of the clause from its underlying position, whereas in Kharia there is no underlying position. Thus, reference to topicalization and left-dislocation is only meant to serve as an analogy and should not be taken literally.

⁵ The prefield encompasses both the “left-detached position” or “LDP” and the “precore slot” or “PrCS” in Role and Reference Grammar. Cf. e.g. Van Valin (2005), especially Chapter 1.

⁶ The postfield corresponds to the “right-detached position” or “RDP” in Role and Reference Grammar (cf. Van Valin, 2005).

Case-syntagmas are obligatory, i.e., syntactically required, thus active and middle signal whether the scenario being portrayed has one, two or—with a few contentive morphemes such as *ter* ‘give’—three arguments. This does not in any way imply that these objects must be overtly mentioned, whether or not their identity is known or recoverable from context.⁷

In sum, the core consists of only the TAM/PERSON-syntagma and the subect marking, while the remainder of the clause consists of optional core-external Case-syntagmas. While the TAM/PERSON-syntagma denotes the action and signals the number of participants involved at the semantic level, a Case-syntagma merely serves to identify these entities, where this is considered necessary.

If this analysis is correct, we should not expect to find ANY pivots involving grammatical operations which apply to the clause above the core level (= the TAM/PERSON-syntagma with its enclitic subject), such as, e.g., coreferential deletion across clauses or reflexivity expressed through syntactic means, as these clause-level but core-external units are not considered grammatical relations such as subject and object here but are more akin to adjuncts. On the other hand, operations which refer to core-level units, such as the enclitic subject, can be subject to such pivots, e.g., “equi-NP deletion” involving complex TAM/PERSON-syntagmas. As we shall see in the following section where the subject is discussed in more detail, these predictions are borne out in Kharia, as there is no evidence whatsoever for a “syntactic subject” in the language which refers to an overt Case-syntagma.⁸

7.1.2 Subject

For convenience, Table 7.1 gives an overview of subject-markers on the TAM/PERSON-syntagma, repeated here from Section 6.2.

⁷ This is stated quite aptly by Hoffmann (1903: 164–165) with respect to Mundari, which appears to be quite similar to Kharia in this respect: “That Mundari should have two differently ending Suffixes for Transitive and Intransitive Predicates [i.e., Active and Middle, JP] is itself very remarkable. *Transitivity and Intransitivity are not so much objective qualities of actions as subjective modes of conceiving actions.*” (my emphasis)

⁸ The absence of syntactic pivots by itself of course does NOT PROVE that the overt complements in Kharia are adjuncts. Nevertheless, if my analysis is correct, then this means that there cannot be any such syntactic pivots in the language, as pivots refer to grammatical relations (or, more precisely, units such as S, A and O), not adjuncts. As I analyze only the enclitic subject marker on the predicate as an overt grammatical relation, this necessarily means that any pivots found can only refer to this unit.

Table 7.1: Person/Number/Honorific marking on the predicate

	<i>Singular</i>	<i>Dual (/HON)</i>		<i>Plural</i>	
		<i>Inclusive</i>	<i>Exclusive</i>	<i>Inclusive</i>	<i>Exclusive</i>
1	= <i>(i)ŋ</i> / = <i>(i)ŋj</i>	= <i>naŋ</i>	= <i>jar</i>	= <i>niŋ</i>	= <i>le</i>
2	= <i>(e)m</i>	= <i>bar</i>		= <i>pe</i>	
3	–	= <i>kiyar</i>		= <i>ki</i> / = <i>may</i>	

The following presents a discussion of the behavior of certain grammatical operations in Kharia which are often considered subject properties in linguistic literature. As we shall see, there is no evidence whatsoever in Kharia for a “syntactic subject” and the only “subject properties”—“equi-NP deletion” and perhaps morphological reflexivization—are both predicate-level (i.e., core-level) properties and do not refer to overt Case-syntagmas.

Coreferential deletion in finite clauses

Coreferential deletion, i.e., the omission of an argument, in this case the “overt subject”, from one clause as it is also contained in the previous clause, is often found to follow an S/A-pivot. Two examples of this are given in the following:

18. *mɔŋ dinu aba=dɔm sou²b=te remakh=o² ro [∅] gam=o²*
 one day father=3POSS all=OBL call=ACT.PST and say=ACT.PST
no “dhãy[=e]=pe.”
 Cmpl hurry=ACT.IRR=2PL
 ‘One day their father called them all and [he] said “Hurry.”’
 [AK, 1:9]
19. *ro be²t=dɔm=ki=jo umay juŋ=o² no i jhãut heke*
 and son=3POSS=PL=ADD NEG.3PL ask=ACT.PST Cmpl what animal QUAL.PRS
lekin ka² kom=ki dho²=ke [∅] mu² go²d=ki=may.
 but bow arrow=PL grab=SEQ emerge C:TEL=Mid.PST=3PL
 ‘And his sons also didn’t ask which animal it is, but, having taken their bows and arrows, [they] went off.’
 [AK, 1:11]

The following examples show, however, that this apparent S/A-pivot is in fact due to topic continuity in coreferential deletion, where the sub-

ject is GENERALLY the topic. Consider the following example, where the subject of the second clause refers to another group of people who carry the Kharia around, not the Kharia themselves, who are the subject of the first clause.

20. *laʔ... khariya lebu=ki pujapaʔh karay=na laʔ=ki=may,*
 then Kharia man=PL sacrifice do=INF IPFV=MID.PST=3PL
soub=aʔ thon, ro ho=ki=te qoli=te qo<ʔb>ko=ke
 all=GEN for and that=PL=OBL palanquin=OBL sit.down-<CAUS>=SEQ
goʔ=ke qoʔq=na laʔ=ki=may.
 carry.on.shoulders =SEQ take=INF IPFV=MID.PST=3PL
 ‘The Kharia men used to perform sacrifices, for all, and [they ≠ Kharia] used to seat them [= the Kharia] on a palanquin and carry them on their shoulders.’ [AK, 2:6]

The following is a similar example, taken from a song. In the second clause, the rain is omitted, although it is not coreferential with the subject of the preceding clause.

21. *koyo=te dhukay=te tapaʔ=te arabquʔ.*
 wind=OBL blow.strongly=ACT.PRS sprinkle=ACT.PRS poor
 ‘The wind is blowing strongly and [the rain] is getting the poor [girl] wet (= is sprinkling the poor one).’ [BB, 4:3]

The following presents an elicited example:

22. *in lebu=ki=te yo=yoʔj ro peʔ nokh=oʔ=ki.*
 1SG person=PL=OBL see=ACT.PST.1SG and rice eat=ACT.PST=PL
 ‘I saw the people and [they] ate rice.’

As these data show, and as was predicted in Section 7.1.1, coreferential deletion in finite clauses in Kharia is not restricted to the “overt subject”

Subject identity with the sequential converb (“conjunctive participle”)

As discussed in 6.6.2.3, the S or A of the sequential converb is generally coreferential with the subject (=S/A) of the main predicate, as in the following example (= the second half of (19) above).

23. ... *ka?* *kom=ki* *dho?*=ke *mu?* *go?**d=ki=may.*
 bow arrow=PL grab=SEQ emerge C:TEL=MID.PST=3PL
 ‘... they took their bows and arrows and set off (i.e., ‘having taken their bows and arrows, they emerged’).’ [AK, 1:11]

As was the case with coreferential deletion involving finite clauses, this is merely a tendency and can probably be accounted for by the tendency to maintain the topic throughout the narrative, which is typically also the subject of the clause. It is, however, only a tendency and counter-examples such as the following, repeated here from 6.6.2.3, are found. For example, an oblique-marked experiencer may be coreferential with the S or A of a converb:

24. *etwa=te* *u=ki=ya?* *haleit* *yo=kon* *lebui* *la?*=ki.
 Etwa=OBL this=PL=GEN condition see=SEQ love EMOT=MID.PST
 ‘Etwa, *seeing* their condition, felt compassion.’ [RD, 2:108]

Similarly, the omitted S or A of the converb may refer to a non-argument (25) or may not be present at all in the sentence (26)-(27):

25. *thor to?* *aw=kon* *jijray=a?* *sori* *daj* *go?**h=o?*.
 few day stay=SEQ porcupine=GEN with send C:TEL=ACT.PST
 ‘[Ø] *having stayed* a few days, [he_i] sent [her_k] off with the porcupine.’
 [HJPa:40]

26. *u=ghay* *gam=kan* *mane* *ho=kar* *raj* *karay=na* *suru*
 this=way say=SEQ umh that=SG.HUM ruling do=INF begin
*go?**h=o?*
 C:TEL=ACT.PST
 ‘[They] *having spoken* thus, umh, he began to rule.’ [MS, 1:185]

27. *tama rāci* *eriya=te* *d̥el=kan* *piṭauriyagarh=jo* *hoy=ki*, ...
 now Ranchi area=OBL come=SEQ Pitauriagarh=ADD become=MID.PST
 ‘Now, [they] *having arrived* in the Ranchi area, they founded Pitauriagarh (= Pitauriagarh also became), [MS, 1:213]

There is thus a clear tendency towards an S/A-pivot with regards to coreferential deletion with the sequential converb, however this is only a tendency and cannot be taken as evidence for a syntactic subject.

Reflexivization

Syntactic reflexivization The reflexive proform *apan* (6.5.1), borrowed from Sadri, generally refers to an agentive subject, as in the following examples:

28. *u khorī boʔ=ki=te lebu=ki ek dusre=te sāghro*
 this village place=PL=OBL person=PL one second=OBL help

karay=na=ʔ ghad, apan sistam oʔj=siʔ=may.
 do=INF=GEN PURP REFL system take.out=PERF=3PL

‘The people of these villages, in order to help one another, have developed (= taken out) their own system.’ [AK, 5:11]

29. *khariya apan ray=te bancay guḍu[ʔ].*
 Kharia REFL culture=OBL save OPT

‘The Kharia should save their own culture.’ [MS, 1:263]

Although *apan* usually follows the “overt subject”, this is not necessarily the case.

30. *apan=aʔ pothi=te ho=kaḥ oʔb-son=oaʔ.*
 REFL=GEN book=OBL that=SG.HUM CAUS-buy=ACT.PST

‘S/he sold his/her own book.’

31. *apan=aʔ pothi=te rayem saroj=te ter=oaʔ.*
 REFL=GEN book=OBL Rayem Saroj=OBL give=ACT.PST

‘Rayem_i gave Saroj_j her_i *_j own book.’

In fact, no “overt subject” is required at all:

32. *aina=te apan=te yo=yoʔj.*
 MIRROR=OBL REFL=OBL see=ACT.PST.1SG

‘I saw myself in the mirror.’

As the following example shows, the reference of the reflexive is not dependent on agency:

33. *apan=aʔ oaʔ=yaʔ telon tay gur goʔd=ki.*
 REFL=GEN house=GEN ROOF ABL fall C.TEL=MID.PST

‘S/he fell from his/her own roof.’

The evidence so far might appear to suggest that syntactic reflexivity is a subject property, where the subject marking on the TAM/PERSON-syntagma is the subject, since no “overt subject” is necessary. The first problem with this, however, is that the reflexive may also refer to an oblique-marked experiencer:

34. *ho=kaṛ=te apan=aʔ ma=qom iyed qel=ki.*
 that=SG.HUM=OBL REFL=GEN mother=3POSS memory come=MID.PST
 ‘S/he remembered his/her own mother.’

However, it may not refer to an oblique-marked recipient:

35. *rayem saroj=te apan=aʔ pothi=te o-y-eṅ=oʔ.*
 Rayem Saroj=OBL REFL=GEN book=OBL CAUS-y-return=ACT.PST
 ‘Rayem_i gave Saroj_j back her, *_j book.’

However, the reflexive may also refer to certain other “oblique” functions, in this case the (non-specified) possessor of the affected body:

36. *apan=aʔ neri=te kosu dhokh=oʔ.*
 REFL=GEN body=OBL pain grab=ACT.PST
 ‘Someone felt pain in their own body.’ (lit. ‘Pain grabbed [his/her/my/. .] own body.’)

We may therefore conclude, as predicted in 7.1.1, that syntactically expressed reflexivity presents no evidence for a “syntactic subject” and is not restricted to ANY notion of “subject”

Morphological reflexivization

The data for morphological reflexivization (6.5.1) are still very meager, hence the following discussion is still somewhat tentative. In all attested cases to date, however, reference is to the enclitic subject of the clause marked on the TAM/PERSON-syntagma:

37. *yo qom=ki=kiyar.*
 see REFL=MID.PST=DU
 ‘They saw themselves (e.g., in the mirror)’

38. *taʔj qom=ki=may komaṅ=te.*
 distribute REFL=MID.PST=3PL meat=OBL
 ‘... they distributed the meat amongst themselves.’ [AK, 1:56]

The following two examples are from Kerkeṭṭā (1990: 12):

39. *koḍṗuru?*=ki *gam*=na=? *hosiyar* *gam* *ḍom*=ta=ki.
 man=PL say=INF=GEN intelligent say REFL=MID.PRS=PL
 ‘Men *call themselves* intelligent in (= of) speech.’

40. *apan dular konselḍug*=a? *mugamte korṗuru?*=ki *juḡhay*
 own love woman=GEN before man=PL much

ḍiṭḡgar ob-yo *ḍom*=ta=ki.
 brave CAUS-see REFL=MID.PRS=PL
 ‘In front of their beloved wives, men *show themselves* [to be] very brave.’

Despite the fact that the evidence for the reflexive use of *ḍom* is so meager, it seems highly likely that reference here will necessarily be to the enclitic subject marked on the TAM/PERSON-syntagma, although this has yet to be demonstrated conclusively.

Relativization

Relativization in Kharia is quite complex and is dealt with in detail in Section 7.6. There it is shown that relativization is possible whether or not the head is a subject in the main or relative clause. In fact, as Table 7.15 at the end of this chapter shows, none of the relativizing strategies are restricted in any way to the subject of the clause, neither the “overt subject” nor the subject marking on the TAM/PERSON-syntagma.

“Equi-NP” deletion

The only pivot in the language which has so far been identified is the S/A-pivot in what is often referred to as “equi-NP” deletion, involving complex predicates denoting *try*, *be able*, *want*, etc. In these cases, the non-overt S or A of the infinitive is obligatorily interpreted as the enclitic subject (= S/A) of the modal auxiliary. Note that this refers only to the enclitic subject marker on the TAM/PERSON-syntagma, not any overt Case-syntagma.

41. *co*=na *lam*=t=*ijn*.
 go=INF want=ACT.PRS=1SG
 ‘I want [I / *you / *he / *she / . .] to go.’

If the S or A of the infinitive is different than the subject of the modal, this must be expressed by a finite complement clause, which usually begins with *no* 'CMPL'

42. *lam=t=ijn* *no* *ho=kaɽ* *co=na*
 want=ACT.PRS=1SG CMPL that=SG.HUM go=MID.IRR
 'I want him/her to go.'

Furthermore, there is no "exceptional case marking" construction in Kharia:

43. **ho=kaɽ=te* / **ho=kaɽ* *co=na* *lam=t=ijn*.
 that=SG.HUM=OBL that=SG.HUM go=INF want=ACT.PRS=1SG
 'I want him/her to go.'

Similar comments hold for *pal* 'be able':

44. *kayom=na* *pal=t=ijn*.
 speak=INF be.able=ACT.PRS=1SG
 'I can [I / *you / *he / *she / . . .] talk.'
45. **ho=kaɽ=te* *kayom=na* *pal=t=ijn*.
 that=SG.HUM=OBL speak=INF be.able=ACT.PRS=1SG
 (but alright with the meaning 'I can talk to him/her.')

The case for *asra* 'hope' is slightly different, although the principle is the same. *asra* takes an infinitive marked for the genitive plus the postposition *thoŋ* 'for'. If the S or A of the infinitive is identical with the subject of *asra*, this is not indicated, as in the following example.

46. *co=na=?* *thoŋ* *asra=t[e]=ijn*.
 go=INF=GEN for hope=ACT.PRS=1SG
 'I hope to go.'

If, however, the S or A of the infinitive is not identical with that of *asra*, this S or A must appear before the infinitive in the genitive. If it does not appear, the infinitive is obligatorily understood as having the same S or A as *asra*.

47. *ho=kaɽ=a?* *co=na=?* *thoŋ* *asra=t[e]=ijn*.
 that=SG.HUM=GEN go=INF=GEN for hope=ACT.PRS=1SG
 'I hope s/he goes.'

Summary of “subject properties”

As the preceding pages show, there is no evidence for a “syntactic subject” in Kharia. The only pivots found in the language, S/A-identity in “equi-NP deletion” and probably with morphological reflexives as well, both refer to the enclitic subject marking on the TAM/PERSON-syntagma, i.e., the core level only, and would seem poor choices for a syntactic subject. On the other hand, all clause-level operations, such as coreferential deletion in finite clauses and with the sequential converb, syntactically expressed reflexivity and relativization, show no signs of functioning along the lines of any pivot. Thus, the term “subject” in Kharia refers only to the enclitic on the TAM/PERSON-syntagma and does not entail any other syntactic properties.

7.1.3 *Object types*

There are two types of “subcategorized” elements in Kharia which are not subjects: (direct) objects and oblique arguments. The object is a well defined category in Kharia: When definite and countable, a Case-syntagma referring to the direct object appears in the oblique case, marked by =*te*. When it is indefinite and/or a mass entity, it usually appears in the direct case. This category corresponds to O or P in typological studies. Furthermore, it may also be “promoted” to the subject in the passive (cf. 6.5.1).

The second group, the oblique arguments, can be divided into two subgroups. The first subgroup consists of units which are generally marked by =*te* if they refer to a recipient (e.g. with *ter* ‘give’) or locative (e.g. with *maṛay* ‘place’), but not consistently in the latter case, although the rules governing this marking are not entirely clear. The second subgroup consists of units marked by postpositions. Included here are beneficiaries, which are generally marked by the postposition *thon* ‘for’, and the intermediate causer in causative forms of underlying transitive predicates or double causatives, which are either marked by the instrumental postposition *buŋ* or by the oblique marker =*te* (cf. 6.3.5).

Elements in these two groups may be considered “arguments” as their presence is assumed either by the underlying semantics, such as the beneficiary with *ter* ‘give’ or the locative with *maṛay* ‘put, place’, or is indicated explicitly on the TAM/PERSON-syntagma itself, e.g., through the benefactive v2 *kay* or (double) causative marking. However, like all other “arguments”, these elements need not be expressed overtly in the clause, nor are they marked for person or number on the TAM/PERSON-syntagma.

Unlike the direct object, oblique arguments may never be “promoted” to subject in the passive. Thus the morpho-syntactic behaviour of a recipient with *ter* ‘give’ is virtually identical to that of a locative adverbial (e.g., with *ḍoko* ‘sit down’): An oblique argument or adjunct may not be promoted to subject in the passive ((49) and (52)), whereas an object may (51).

48. Active: *ho=kaṛ* *u* *ṭhāro=te* *ḍoko=ki*.
 that=SG.HUM this place=OBL sit.down=MID.PST
 ‘S/he sat down in this place.’
49. Passive: **u* *ṭhāro* *ḍoko* *ḍom=ki*. ‘This place was sat on.’
 this place sit.down PASS=MID.PST
50. Active: *ho=kaṛ* *am=te* *poṭhi=te* *ter=or?*
 that=SG.HUM 2SG=OBL book=OBL give=ACT.PST
 ‘S/he gave you the book.’
51. Passive: *am=te* *poṭhi* *ter* *ḍom=ki*. ‘A book was given to you.’
 2SG=OBL book give PASS=MID.PST
52. Passive: **am* *poṭhi=te* *ter* *ḍom=ki=m*. ‘You were given the book.’
 2SG book=OBL give PASS=MID.PST=2SG

Finally, in double causatives or in the causative of underlying transitive predicates, only the S/O (or S/P) of the underlying predicate (i.e., theme or patient) may be promoted to the subject in the passive, never the intermediate causer. Thus, although the intermediate causer can be marked for the oblique case, it is not an object but merely an oblique argument.

7.2 INTERROGATIVES

The present section deals with different types of interrogatives, beginning with polar questions (7.2.1). This is followed by a discussion of constituent or “*wh-*” questions (7.2.2) with a brief sub-section dedicated to multiple constituent questions (7.2.2.1). Finally, 7.2.3 deals with alternative questions.

7.2.1 *Polar questions*

There are two main strategies found with polar questions:

⇒ The polar question has the same structure as the declarative, but differs with respect to intonation, rising towards the end of the utterance, before it falls off (cf. Rehberg, 2003: 38f., 93). This is demonstrated in the following two examples, which are only marked as questions by intonation:

53. *tay konon bhai=te juṅ=o?* *no* “*beghma, am*
 then small brother=OBL ask=ACT.PST CMLP *Beghma 2sg*
ḍa? uḥ=oʔb?”
 water drink=ACT.PST.2SG
 ‘Then he asked his younger brother “*Beghma, did you drink water?*”’
 [AK, 1:31]

54. *poṛoṅ berod=ki ro gam=o?* *no* “*iyar, am=te*
 hare arise=MID.PST and say=ACT.PST CMLP *friend 2sg=OBL*
lemed um la?=ta?”
 sleep NEG EMOT=MID.PRS
 ‘The hare arose and said “*Friend, aren’t you tired?*”’ [BB, 3:21]

⇒ The second means of forming polar questions is with either the interrogative particle *i* ‘*q*’, which also means ‘what?’, or *no* ‘*q*’, which is homophonous with the complementizer and also *no* ‘*or*’, which probably all have the same origin.⁹ Either of these may appear at the beginning or, more commonly, at the end of the question (but preceding a vocative element).

55. *o? ḍiyar=na deri bij=ḍom=te juṅ=o?*
 house enter=INF while daughter=3POSS=OBL ask=ACT.PST
 “*gach=ob no, beṭi?*”
 fry=ACT.PST.2SG *q* daughter
 ‘While entering the house, she asked her daughter “*Did you fry him, daughter?*”’
 [BB, 1:79]

⁹ Cf., e.g., Hindi *ki* 1. ‘*q*’; 2. ‘*or*’; 3. ‘*CMLP*’

56. *kinbhar=te munga daru on=oʔb no, mā joy?*
 courtyard=OBL *munga* tree plant=ACT.PST.2SG Q mother VOC
 ‘Did you plant the *munga* tree in the courtyard, oh Mother?’
 [TK, 4:1]

57. *hontay ho=ki ho=kaʔ=te juŋ=oʔ=may: “laʔ am ber*
 then that=PL that=SG.HUM=OBL ask=ACT.PST=3PL then 2SG who
heke=m? i am eliyah heke=m?”
 QUAL.PRS=2SG Q 2SG Elijah QUAL.PRS=2SG
 ‘Then they asked him: “Then who are you? *Are you Elijah?*”’
 [HJPa:87, ln. 21]

Very rarely, probably for “poetic effect”, the interrogative may appear elsewhere between two phonological words. The following, from a traditional song, is one of only two examples contained in our data.

58. *pap rupi samudar=te dube no*
 sin having.the.attribute.of ocean=OBL drown(ITR) Q
goʔdʔ=ta=ijn?
 C:TEL= MID.PRS=1SG
 ‘Will I drown in the sinful ocean?’ [SaK/RoD, 1:4]

7.2.2 Constituent (“wh-”) questions

Constituent questions are formed with interrogative proforms which usually appear directly before the TAM/PERSON-syntagma, the unmarked focus position.

59. *agar botoŋ=na=niŋ laʔ odoʔ atu eŋ=na=niŋ?*
 if fear=MID.IRR=1PL.INCL then again where return=MID.IRR=1PL.INCL
 ‘If we get afraid, then *where* will we return to again?’ [MS, 1:35]
60. *tobdaʔ daʔ=te i=ghay uqʔ=e=niŋ?*
 mud water=OBL what=way drink=ACT.IRR=1PL.INCL
 ‘How will we drink muddy water?’ [MS, 1:126]
61. *laʔ tama rajj ber calay[=e]?*
 then now kingdom who drive=ACT.IRR
 ‘Now *who* will lead the kingdom?’ [MS, 1:180]

Despite the tendency to place the interrogative directly before the TAM/PERSON-syntagma, this is not obligatory, as the following example shows. Note that *swarag* ‘heaven’, which intervenes between the question word and the TAM/PERSON-syntagma, is also focal in this example.

62. *ho khorej=gā moŋ kauwa qel=ki ro konthed=ki=te*
 that time=FOC one crow come=MID.PST and bird=PL=OBL

gam=te no “am=pe u papi lebu=ya? kunḍu?=te
 say=ACT.PRS Cmpl 2=2PL this sinful man=GEN child=OBL

ina swarag qor=tej=pe?”
 why heaven take=ACT.PROG=2PL

‘At that very moment a crow came and said to the birds “*Why* are you taking this sinful man’s child to heaven?”’ [BB, 2:29]

The one position that is only marginally acceptable for the interrogative element is when the interrogative follows the TAM/PERSON-syntagma, as all speakers only grudgingly accepted examples such as the following.

63. *qel=ki behar? ‘Who came?’*
 come=MID.PST who

Comments on such examples ranged from “not that good” to “a little strange” Nonetheless, only very few outrightly rejected such examples, despite the fact that they found them so odd.

Finally, recall from Chapter 4 that interrogatives may also function as predicates, with no derivational morphology:

64. *la? etna qisa? kaŋ go^oq=ki=niŋ je*
 then so.much far get.lost C:TEL=MID.PST=1PL.INCL CREL

i=għay=na?

what=way=MID.IRR

‘We have been lost for such a long distance that we won’t find our way (= *how will it become?*)’ [MS, 1:130]

7.2.2.1 Multiple constituent questions

Except for the general disliking of post-predicative interrogative elements mentioned in the last section, there seems to be no fixed order of

interrogative elements in multiple constituent questions. The following presents a few examples.

65. *ber am=bar=te i ter=o?* ? / *am=bar=te ber i ter=o?* ?
 who 2=2DU=OBL what give=ACT.PST

ber i ter=o? am=bar=te?
 'Who gave you what?'

On the other hand, the following permutations received responses ranging from "not so good" to "a little strange" or even "maybe it's O.K." At any rate, whether or not these forms are "really" grammatical to most speakers in non-interview situations, they clearly demonstrate a strong preference for non-post-predicative interrogatives.

66. *(?) i am=bar=te ter=o? ber?* / *(?) am=bar=te i ter=o? ber?*

67. *(?) ber am=bar=te ter=o? i?* / *(?) am=bar=te ber ter=o? i?*

Questions with three interrogatives were also accepted in any order, as long as the interrogative element did not occur post-predicatively:

68. *ber ber=te i ter=o?* ? 'Who gave who what?'
 who who=OBL what give=ACT.PST

69. *ber i ber=te ter=o?* ?

70. *i ber ber=te ter=o?* ? (found by one speaker to be "a little strange")

71. *i ber=te ber ter=o?* ? (found by the same one speaker to be "a little strange")

72. *ber=te i ber ter=o?* ?

73. *ber=te ber i ter=o?* ?

Despite the tendency just mentioned to avoid post-predicative interrogatives, there are two examples in the corpus where speakers did not object to this. The reasons for this are not entirely clear, although it may be

due to the fact that both interrogatives have identical meanings.¹⁰ For example, all of the four speakers questioned accepted all of the following four permutations without reservation. Further research is necessary on this topic.

74. *behar ber=te gil=o?* / *ber=te behar gil=o?* ? ‘Who hit who?’
 who who=OBL hit=ACT.PST

75. *ber gil=o?* *ber=te?* / *ber=te gil=o?* *ber?* ‘Who hit who?’
 who hit=ACT.PST who=OBL

7.2.3 Alternative questions

Alternative questions are formed simply by the use of *no* ‘q; or’ between the two alternatives:

76. *raksin khis buṅ “kongher=te tama no lo?dho nog[=e]=iṅ?”*
 witch anger INST boy=OBL now q later eat=ACT.IRR=1SG
lekhe hoy=taj.”
 like become=MID.PROG
 ‘The witch angrily thinks (= she is becoming through anger like)
 “Shall I eat the boy *now or later?*”’ [BB, 1:47]

Another common strategy is to place the expression *numbo?* (< *no+umbo?* ‘or not’) at the end of an otherwise declarative statement. Alternatively, as Malhotra (1982: 281) notes, the TAM/PERSON-syntagma is repeated in its negative form:

77. *ho=kaṛ tuḍa de=na no um de=na?*
 that=SG.HUM tomorrow come=MID.IRR q NEG come=MID.IRR
 ‘Will he come tomorrow *or not?*’

¹⁰ There is no difference in meaning between *behar* and *ber*, and both mean ‘who?’, regardless of their function in the clause (i.e., subject, object, etc.).

7.3 QUALITATIVE PREDICATION—“NOMINAL” AND “COPULAR” SENTENCES

7.3.1 *General introduction*

The basic structure of qualitative predicates in Kharia is given in Table 7.2, where “X” represents any kind of Case-syntagma. The qualitative marker (QUAL) may appear either before or after the semantic base “X”, but may not occupy both slots simultaneously.

Table 7.2: The schematic structure of the qualitative predicate in Kharia

(QUAL) X (QUAL)

The semantic base of the qualitative predicate is formed as described in Chapter 5. The qualitative marker is often “lacking” with a stative, present or atemporal interpretation, but with a non-present or non-stative interpretation, the qualitative marker is obligatory. A further distinction is made—in all TAM categories—between stative and dynamic qualitative predication. The following comments pertain to stative qualitative predication only. Dynamic qualitative predication will be dealt with in Section 7.3.4.

There are two types of stative qualitative predication which are distinguished morphosyntactically in Kharia, but only in the present tense. Outside of the present, this distinction is neutralized. In the first type, which will be referred to here as “NON-INHERENT qualitative predication”, the position of an entity is given or a statement is made about a temporary or non-essential characteristic of this entity. The second type is “INHERENT qualitative predication”, in which an inherent property of an entity is depicted (e.g. *I am a man, That is a tree*). “Inherent” should be taken here as meaning something akin to “central or most important characteristic” and will undoubtedly be subject at least to some extent to a particular speaker’s judgement.

To demonstrate the difference between these two types of qualitative predication, consider the following examples:

- | <i>Non-Inherent</i> | <i>Inherent</i> |
|---|----------------------------------|
| 78. a. <i>ho=kaɾ hosiyaɾ ayi²j</i> | b. <i>ho=kaɾ hosiyaɾ heke</i> |
| that=SG.HUM intelligent QUAL.PRS | that=SG.HUM intelligent QUAL.PRS |
| ‘She is intelligent.’ | ‘She is the intelligent one.’ |

In the first example, a statement is being made about a certain woman and we are being told that she is intelligent. While this may be important, the use of the non-inherent qualitative marker *ayiʔj* denotes that it is not the defining characteristic of this person but merely one of many (e.g., *human, female, mother, middle-aged, intelligent, . . .*). On the other hand, the use of the inherent qualitative marker *heke* in the second example is most natural in a conversation about an intelligent person unknown to one of the speech participants. When this same woman turns up at the scene of the conversation, the second example above could be used to identify her. I.e., in this context, only the fact that she is intelligent (or rather: “the intelligent one”) is of interest.¹¹

TAM-values for both types of qualitative predication are expressed by means of what are referred to in a number of previous studies (e.g., Peterson, 2007, 2008) as the copula, which has a number of suppletive forms for the present tense. Unlike the TAM/basic voice markers of narrative predication, these forms are phonologically independent words and may appear either before or after the semantic base of the predicate.

Since these earlier publications it has become clear that there are convincing reasons for not considering these forms to be copulas in the true sense of the word (see next section), hence these forms will be collectively referred to in the following simply as qualitative predicative markers (QUAL). In the following section, the status of these markers is clarified and arguments against considering them copulas are given, before proceeding in 7.3.3 and 7.3.4 to discuss their forms. Finally, in 7.3.5 it is shown that the presence or absence of these qualitative markers in stative qualitative predication is semantically relevant.

7.3.2 *Does Kharia have a copula?*

Following general convention, a copula is considered here to be a form—usually a verb form—which possesses the following characteristics:

- In the case of a verbal copula, the presence of this element is required for language-specific reasons, most notably as the particular language requires EVERY PREDICATE TO CONTAIN A VERBAL ELEMENT, in this case a “dummy verb” which serves to indicate tense, aspect and mood.

¹¹ Thus, the difference between “inherent” and “non-inherent” appears to differ somewhat from the “individual-” vs. “stage-level” distinction, although the two are certainly closely related.

- A copula is an OBLIGATORY element found in qualitative predication and contributes very little semantically to the predicate, if at all.
- With respect to optionality and “zero copulas”: Although ellipsis is a common phenomenon in most languages, only if it can be shown that a language generally requires a copula but that this is “dropped” in ellipsis WITH NO RESULTANT SEMANTIC DIFFERENCES are we justified in assuming a “zero copula” or that the copula has been “dropped”

Let us now turn to each of these points in more detail. In doing so, we will closely adhere to the argumentation found in Benveniste (1966).

With regard to the first point mentioned above: A “copula” in those languages which clearly possess this category, e.g., the modern Western European languages such as English, German or French, is required as every predicate in these languages must contain a verbal element expressing TAM and person marking. As was shown in Chapter 4, however, Kharia is most easily described without recourse to categories such as “noun” and “verb” Instead, we merely find TAM/PERSON-syntagmas and Case-syntagmas. As there is no such thing as a nominal sentence in Kharia, the presence of a verbal copula, needed to fulfill the language-specific requirement that every predicate contains a verb, does not hold here.

The second point mentioned above, that of obligatoriness, is perhaps the most important of the three: As Benveniste (1966) argues, it is important to distinguish among languages which at first glance would appear to have a copula between those which REQUIRE this element and those where this element is STRUCTURALLY OPTIONAL: Whereas the term “copula” would appear justified in the first type (e.g., English, French, German), this is quite different in the second type. In these languages, it is quite common practice to assume the presence of a “zero copula” in an ellipsis for those cases where, from an English-speaking perspective, the copula appears to be missing. However, what this implies is that such a form SHOULD be present but has been “left out” in a particular utterance, thereby assuming *a priori* that the language in question must have a copula.

This immediately brings us to the third point mentioned above: As Benveniste (1966) shows for a number of older Indo-European languages (especially Greek), and as will be shown below in 7.3.5 for Kharia, the presence or absence of such a form can express a semantic difference between what we may loosely term “narrative description” (in the presence of a qualitative marker) and “general or self-evident truth” (in the absence of such a form). As such, we are not justified here in assuming

that a form of the “copula” has been “dropped” Rather, the absence of such a form is itself semantically relevant.

Summarizing, as the term “copula (verb)” is generally taken to refer to an obligatory “dummy” element required in qualitative predication due to a language-specific requirement that all predicates contain a verb, the use of this term cannot be justified in the description of a language such as Kharia, where the presumed “copula” is not obligatory (with a general, atemporal interpretation) and where there is no compelling evidence for the presence of nouns and verbs. Instead, those forms referred to in this author’s previous works on Kharia as copulas will be considered markers of qualitative predication: When present, these forms ground the utterance in terms of tense, aspect and mood and denote whether the predicate is to be interpreted as stative (‘be’) or dynamic (‘become’). When these forms are not present in a qualitative predicate, the utterance is a “general or self-evident truth”

7.3.3 *Stative qualitative predication—the forms*

As discussed above, in the present-tense a distinction is made between INHERENT and NON-INHERENT qualitative predication, both of which have suppletive positive and negative forms. Table 7.3 presents the present-tense forms of these markers. These forms do not mark for basic voice but do mark for PERS/NUM/HON.

Both the non-negated form *heke* and the negated form *nalage* have been borrowed directly from Sadri. *ayiʔj(d)* on the other hand would also appear to be an Indo-Aryan loan word, although it is not entirely clear from which language it has been borrowed. It is clearly not from (modern) Sadri, where the corresponding form is *āh-*, however the presence of the 3rd person, non-honorific form *əich* ‘he/she/it is’ in Maithili (Yadav, 1996: 217ff.) would seem to indicate that it has been borrowed from some Indo-Aryan source, and the stem *ach-* is also found elsewhere in eastern Indo-Aryan, such as Bengali. Instead of *ayiʔj(d)*, the form *ārij* is occasionally found in older texts.

Table 7.3: The present-tense qualitative markers

<i>Type of stative qualitative predication</i>	<i>Non-negated form</i>	<i>Negated form</i>
Inherent qualities	<i>heke</i>	<i>nalage</i>
Non-inherent qualities	<i>ayiʔj(d)</i>	<i>umboŋʔj(d)</i>

Tables 7.4 and 7.5 show the forms of the non-negated, inherent and non-inherent qualitative markers *heke* and *ayi²j(d)*, respectively: *nalage* conjugates similarly to *heke* and *umboŕi²j(d)* (with the rare alternate forms *uboŕej(d)* and *umboŕej(d)*) similarly to *ayi²j(d)*.¹²

Table 7.4: The present-tense forms of the non-negated, inherent qualitative marker

	Singular	Dual/ ^{HON}		Plural	
		Incl.	Excl.	Incl.	Excl.
1st person	<i>hek=ijn</i>	<i>heke=naŕ</i>	<i>heke=jar</i>	<i>heke=niŕ</i>	<i>heke=le</i>
2nd person	<i>hek=em</i>	<i>heke=bar</i>		<i>heke=pe</i>	
3rd person	<i>heke</i>	<i>heke=kiyar</i>		<i>heke=ki / heke=may</i> <i>hekãy</i> ¹³	

Table 7.5: The present-tense forms of the non-negated, non-inherent qualitative marker

	Singular	Dual/ ^{HON}		Plural	
		Incl.	Excl.	Incl.	Excl.
1st person	<i>ayi²jd=ijn</i>	<i>ayi²j=naŕ</i>	<i>ayi²j=jar</i>	<i>ayi²j=niŕ</i>	<i>ayi²j=le</i>
2nd person	<i>ayi²jd=em</i>	<i>ayi²j=bar</i>		<i>ayi²j=pe</i>	
3rd person	<i>ayi²j</i>	<i>ayi²j=kiyar</i>		<i>ayi²j=ki / ayi²j=may</i>	

Unlike the present tense in dynamic narrative predication, the present-tense forms of these stative qualitative markers refer only to the here and now.¹⁴ For expressing stative qualitative predication with a non-present or non-indicative interpretation, the morpheme *aw* is used, which otherwise has the meaning ‘stay, remain; live’, and the inherent/non-inherent distinction is neutralized. Also, unlike the forms with a present-tense interpretation, *aw* as a marker of stative qualitative predication is always

¹² Note that the final *-d* in *ayi²j(d)* and *umboŕi²j(d)* is omitted before all person markers other than those of the first- and second-persons, singular, the only two person markers which begin with a vowel. This is due to the strict (C)V(C) syllable structure of the native vocabulary (cf. 2.3–2.4).

¹³ The form *hekãy* is the Sadri form of the 3rd person, plural. This form is occasionally found instead of the “regular” forms *hekeki* and *hekemay*.

¹⁴ They may perhaps also be used with a “historical present” interpretation, although no examples for this use have yet come to my attention.

marked explicitly for the middle voice. The following provides a few examples of stative qualitative predication.

Inherent qualities—*heke and nalage*

79. *u kahani u goṭa duniya=te lebu=ki=ya?* *kahani heke.*
 this story this entire world=OBL person=PL=GEN story QUAL.PRS
 ‘This *is* the story of the people on the entire world.’ [AK, 3:1]

80. *muda tam jou lebu=ki umay koṅ=sikh=o?* *no*
 but now up.to person=PL NEG.3PL know=PERF=ACT.PST CMPL
u=je? *i daru heke.*
 this=SG.NHUM what tree QUAL.PRS
 ‘But to this day, people have not found out what tree it *is*.’
 [BB, 1:18]

81. ... *no je khajar tar=sikh=o?=may ho=kaṭ=a?*
 and CREL deer kill=PERF=ACT.PST=3PL that=SG.HUM=GEN
komay=ko nalage...
 meat=CNTR QUAL.NEG.PRS
 and it *isn't* the meat of the deer that they had killed. . .’
 [AK, 1:59]

As noted above, the qualitative markers may also precede a Case-syntagma. This is especially common with “heavy” semantic bases (82) or when the semantic base is focused or contrasted ((87) below):

82. *u=je[?]=ga heke kharīya gotar=a?* *choṭka kahani.*
 this=SG.NHUM=FOC QUAL.PRS Kharia clan=GEN small story
 ‘This *is* a short history of the Kharia clan.’ [AK, 1:70]

Recall from 6.7.1 that Malhotra (1982: 286) cites a few examples in which the inherent present qualitative marker is negated similarly to dynamic narrative predicates, i.e., *um heke* instead of *nalage*. This usage is quite rare, however, and the form *nalage* is much more common and is considered standard.

Non-inherent qualities—ayiʔj(d) and umboʔiʔj(d)

83. *ho=kiyar=yaʔ tomoq=te dāraʔ ayiʔj.*
 that=DU=GEN mouth=OBL stick QUAL.PRS
 ‘There *is* a stick in their mouths.’ [TK, 1:44]

84. *u khorī=te beṭa umboʔiʔj=may.*
 this village.section=OBL boy QUAL.NEG.PRS=3PL
 ‘In this village section there *aren’t* [any] boys.’ [RK, 5:3]

85. *ḍairgāw ro khariyaṭoli hinte tama khariya=ki umboʔiʔj=ki.*
 Dayrgaon and Khariatoli LOC now Kharia=PL QUAL.NEG.PRS =PL
 ‘There *aren’t* [any] Kharia now in Dairgaon and Khariatoli.’
 [MT, 1:231]

As noted in 5.2, the non-inherent qualitative marker is also used to denote predicative possession:

86. *muʔjḍaʔ gam=oʔ, “umboʔ bhai, iṅ=aʔ=te samay*
 ant say=ACT.PST no brother 1SG=GEN=OBL time
umboʔiʔj.”
 QUAL.NEG.PRS
 ‘The ant said, “No brother, I *have no* time.”’ [TK, 2:25]

Finally, as discussed in Section 6.8.2.4, these forms are also used in conjunction with the infinitive to denote obligation or necessity.

Non-present stative qualitative predication with aw

87. *raja=yaʔ jimi aw=ki sembho oḍo rani=yaʔ jimi*
 king=GEN name QUAL=MID.PST Sembho and queen=GEN name
aw=ki ḍakay.
 QUAL=MID.PST Dakay
 ‘The king’s name *was* Sembho and the queen’s name *was* Dakay.’
 [AK, 1:4]

As the following example shows, *aw* is used not only for past or future states but also for habitual and iterative situations:

88. *jhari po?da hinte mu?u la? mu?u ga?ysi*
 all village LOC one.HUM then one.HUM spy

aw=ta=ki=ga.

QUAL=MID.PRS=PL=FOC

‘In all villages there *are* [always] some spies.’ [MT, 1:207]

7.3.4 Dynamic qualitative predication

In addition to stative qualitative predication, Kharia also possesses forms denoting dynamic qualitative predication, corresponding roughly to English ‘become’, although the translation ‘take place’ is often more appropriate. These forms—in all TAM categories—derive from the morpheme *hoy*, which has been borrowed from Sadri or, less commonly, *bone*, also from Sadri. These forms will be glossed as ‘become’ for convenience.

89. *sou?b se maha=ya? biha hoy=si?*
 all ABL big=GEN wedding become=PERF

‘The oldest [sister] has married (= the wedding of the biggest *has become*).’ [AK, 4:9]

90. *u jharkhand? alag hoy=ki ho bhere tay*
 this Jharkhand separate become=MID.PST that time ABL

gha? iskul ko?ghel=jo ho=ki khariya kayam=ta=may.

therefore school vicinity=ADD that=PL Kharia speak=MID.PRS=3PL

‘Since the time Jharkhand *became* independent, therefore they speak Kharia in school as well.’ [AK, 5:24]

91. *ka?i?j deri=te rel chu?te=na=ya? bera hoy=ki.*
 some time=OBL train leave=INF=GEN time become=MID.PST

‘After some time, the time for the train to depart arrived (= it *became* the time. .).’ [RD, 1:10]

Very rarely, and generally in the language of the southern speakers, the form *hoy* is found in environments where one of the stative qualitative markers would be expected.

92. *sow-d?ay=d?om=a[?]* hoy=ki *jhariyo.*
 spouse-woman=3POSS=GEN become=MID.PST Jhariyo

‘His wife’s [name] *was* Jhariyo.’ [MS, 1:19]

7.3.5 “Zero copulas”?

In his discussion, Benveniste (1966) differentiates between qualitative predicates¹⁵ consisting only of a nominal phrase and those consisting of a nominal phrase plus the verb ‘be’ (*être*). Applied to Kharia, this equates to that part of the predicate with the form of a Case-syntagma (Benveniste’s *phrase nominale*) and the stative qualitative predicate marker (*le verb «être»*). As Benveniste (1966: 162) notes for his study of these constructions in various Indo-European languages, especially Greek, the presence or absence of the verb ‘be’ is not haphazard but can be summarized as follows:

- The absence of the verb ‘be’
 - is always connected to direct discourse and
 - always serves for assertions which express a “general truth”
- On the other hand, the presence of the verb ‘be’ is linked to
 - the narration of a fact and
 - the description of a manner of being or of a situation

In other words, the absence of ‘be’ states the obvious, reminds the listener of a general truth, or cites something (being portrayed as self-evident) as proof, while the presence of ‘be’ informs the listener and gives the particulars of a situation.

As such, the difference between the presence or absence of the verb ‘be’ is directly related to discourse registers (Benveniste, 1966: 165)¹⁶—direct discourse as opposed to narration—as well as to whether what is being predicated is beyond any doubt, such as a general truth (or what the speaker wishes to portray as an unquestionable truth). This fits in well with the fact that the bare nominal phrase is not marked for tense, aspect or mood: As the information denoted by the predicate is considered beyond question and an obvious fact, it need not be “grounded” or

¹⁵ Note that Benveniste does not speak of qualitative predication but rather “the verbal function” (*la fonction verbale*) to refer to the semantic function of predication in general, as opposed to verbal FORMS. However, as he is clearly dealing with what I term stative qualitative predication, this is only a terminological issue and does not affect our argumentation here.

¹⁶ For example, as Benveniste shows, while the presence of the verb ‘be’ (Greek *ἔστι*) is quite common in the works of Herodotus, a historian narrating scenes and events, the use of this form is considerably less, e.g., in the epics, where stretches of discourse are also found.

situated with respect to tense, aspect or mood. It is essentially atemporal—or at the very least valid for the time of reference—and indicative, and aspectual distinctions are irrelevant.

The criteria Benveniste (1966) proposes for these languages fit remarkably well with the Kharia data. To begin with, we note here that the presence of forms such as *heke* and *ayiʔj* are quite common in our corpus, as these texts are almost entirely narratives. Hence, the use of these forms may be considered unmarked for the narrative register. Especially common is the use of these forms at the beginning or end of a narrative:

Beginning of a narrative: (= (79) above)

93. *u kahani u goṭa duniya=te lebu=ki=yaʔ kahani heke.*
 this story this entire world=OBL person=PL=GEN story QUAL.PRS
 ‘This *is* the story of the people on this entire world.’ [AK, 3:1]

End of a narrative:

94. *u=jeʔ=ga heke manus jati=yaʔ kahani.*
 this=SG.NHUM=FOC QUAL.PRS man ethnic.group=GEN story
 ‘This *is* the story of humanity.’ [AK, 3:17–18]

95. *ro mane=te=may no u beṭa ro beṭi=kiyar=ga*
 and believe=ACT.PRS=3PL CMPL this boy and girl=DU=FOC
loʔqho mugam col=ke soub duniya=yaʔ lebu=ki=yaʔ
 after forward go=SEQ all world=GEN person=PL=GEN
ayo ro aba heke=kiyar.
 mother and father QUAL.PRS=DU
 ‘And they believe that this boy and girl later, having gone forth, *are* the mother and father of the people of the whole world.’

96. *u=jeʔ=ga el=aʔ kuṭum oḍo[ʔ]... pariwar=aʔ*
 this=SG.NHUM=FOC 1PL.EXCL=GEN family and family=GEN
choṭka-san parice heke.
 small-APPROX introduction QUAL.PRS
 ‘This *is* a shortish introduction to our family and relatives.’ [AK, 4:15]

As one of the primary functions—perhaps the primary function—of narration is to inform the listeners of events and situations, the presence of these stative qualitative markers in narration is to be expected. As the following examples show, this is also true of information within the narrative itself, not only at its beginning or end:

[AK, 1:9–11]

‘One day their father called them all and said “Hurry, kill and bring back this kind of animal from the forest”, and he hurries off.’

97. *aba=qom=kiyar um=kiyar batay=o?* *no i jinis*
 father=3POSS=HON NEG=HON tell=ACT.PST Cmpl what animal

heke, i jhāut heke lekin, muda, qaṇ
 QUAL.PRS what animal QUAL.PRS but but send

goṭh=o? sou²b=te=ga.

C:TEL=ACT.PST all=OBL=FOC

‘Their father didn’t tell them which animal it *is*, which animal it *is*, but he sent them all off.’

98. *ro be²t=qom=ki=jo umay juṇ=o?* *no i jhāut heke*
 and son=3POSS=PL=ADD NEG.3PL ask=ACT.PST Cmpl what animal QUAL.PRS

lekin ka? kom=ki dho?=ke mu? go²d=ki=may.

but bow artow=PL grab=SEQ emerge C:TEL=MID.PST=3PL

‘And his sons also didn’t ask which animal it *is*, but, having taken their bows and arrows, they went off.’

There are also cases, such as the following, in which there is indeed evidence that a form of ‘be’ has been “dropped” in an ellipsis. This is seen by the fact that *heke* is used in this example once for the older son, *rohit*, and is not repeated for the second son, *jerab*. As we will see below, this is quite different from those cases in which ‘be’ is not found in qualitative predication but where it is also not elliptical.

99. *maha-e=qom=a?* *jimi heke rohit, oḍo? konon=a?* *Jerab,*
 big-??=3POSS=GEN name QUAL.PRS Rohit and small=GEN Jerab

oḍo[?] *ho=kaṛ=a?* *tay konon bahin=kiyar tama*

and that=SG.HUM=GEN ABL small sister=HON now

hoʔ=te=ga a[yiʔ]j=kiyar kolej karay=te=kiyar, simqɑʔ=te=ga
 house=OBL=FOC QUAL.PRS=HON college do=ACT.PRS=HON Simdega=OBL=FOC
 ‘Her older [son]’s name *is* Rohit, and the younger one’s [*is*] Jerab,
 and her younger sister (HON) *is* now at home, [and] goes to college,
 in Simdega.’ [AK, 4:11]

As the second line of this last example shows, the comments given here with respect to narration and ‘be’ apply equally to both *heke* and *ayiʔj*. Here a few more examples for *ayiʔj*:

Beginning of a narrative:

100. *tama in anin=aʔ gotar=aʔ kahni batay=na*
 now 1SG 1PL.INCL=GEN clan=GEN story tell=INF
col=taʔjq=in.
 GO=MID.PROG=1SG
 ‘Now I am going to tell the story of our clan.’ [AK, 1:1–2]

101. *kahani lebu=ki khoʔi=ki=te kayom=ta=ki. u=ghay*
 story person=PL village.section=PL=OBL speak=MID.PRS=PL this=way
ayiʔj.
 QUAL.PRS
 ‘The people tell [this] story in the villages. It goes (= *is*) like this.’

The narration of information:

102. *simra batay=oʔ “dhāy[=e] hēʔqō, ikon, hēʔqō u rochoʔb*
 Simra tell=ACT.PST hurry=ACT.IRR here umh here this side
qɑʔ ayiʔj.”
 water QUAL.PRS
 ‘Simra said “Hurry! Here, umh, over here *there is* water.”’
 [AK, 1:33]

“Nominal” predicates—qualitative predication without ‘be’

We now turn to examples of qualitative predication in which there is no element which translates as ‘be’. The following presents an example for a kind of “general truth”. In this example, the Kharia are migrating throughout India on their way to their present-day homeland and have

been living for some time at the confluence of the Yamuna and Ganges rivers. At this point in the story the speakers are expressing that they feel relatively safe where they now are, as those who have been pursuing them will have a difficult time crossing the rivers, due to the fact that they have swollen and are now very large:

103. *samudar lekhe ompay=ki boṛe boṛe. la? hote paro=na*
 ocean like river=PL big REP then here cross=INF
muskil heke.
 difficult QUAL.PRS
 ‘The rivers [are] very big, like an ocean. Therefore (= then) it is difficult to cross here.’ [MS, 1:157]

In the first half of this example of direct discourse, where there is no unit translating as ‘be’, the speakers are stating a fact that is obvious to the others, i.e., that the rivers at the place where the interlocutors are located are full of water, like an ocean. What is being narrated here is in the second half of this utterance, when the speakers state that those following the Kharia (not mentioned here directly) will not be able to cross the rivers here because of this, i.e., it will be difficult to cross the rivers here (narrated information), for the obvious reason that the rivers are full (self-evident truth as background information).

The following example, a few lines earlier from the same story [MS, 1:141], is in reference to how the Kharia got to where they are in example (103) above. Here, the speakers are referring to the immediate surroundings of the wandering Kharia, which are known to all present in the situation, that is, the speakers in the following are merely stating the obvious and are not in any way informing their interlocutors of their present situation:

104. *u=tiʔj tay lebu del=taʔj=ki ro u=tiʔj=jo*
 this=side ABL person come=MID.PROG=PL and this=side=ADD
jamuna u=tiʔj ganḡa maha ompay.
 Yamuna this=side Ganges great river
 ‘People are coming from this direction and [on] this side [is] the Yamuna [and on] this side [is] the mighty Ganges River.’

Again, this self-evident fact is being stated as background argumentation for the next line of text, which reads (in English) “How will we cross over? Oh God!” (MS, 1:142). As the Kharia at this point of the story are

being pursued by their enemy, what is of primary concern here is how to escape, not that the rivers are there, which is obvious in this context.

The following example is very similar in structure. In this episode [MS, 1:37], the Kharia are wandering through the Himalayas on their way to their present-day homeland. In the first half of this example, the speaker—again in direct discourse—is reminding his listeners of their position in the Himalayas, which is known to all who are wandering. As in (103), what is being narrated here is the fact that it will be difficult to walk between the Khyber Pass and Tibet for the (obvious) reason that these are below the Himalayas:

105. *himalay* *biru=wa?* *tuta* *ti?j* *khaybar* *na*
 Himalaya mountain=GEN bottom side Khyber.Pass and
tibat, *la?* *mo?jhi=te* *co=na=ko*
 Tibet then middle=OBL go=INF=CNTR
moskil *hoy=na.*
 difficult become=MID.IRR
 ‘Below the Himalayan mountains [*are*] the Khyber Pass and Tibet, then and *it will be* difficult to walk between them.’

Finally, one last example which again demonstrates that no form of ‘be’ is found where a general truth is being stated, although this is not a case of direct discourse:

[MT, 1:64]

106. *akil=a?* *kayom* *no* *musni?* *a=ti?j* *tay=ko* *qel*
 mind=GEN matter Cmpl one.day Q=side ABL=CNTR come
hamba?=ki *mo?* *maha* *dano* *ro* *po?da=ya?* *jha?*
 SUD=MID.PST one big demon and village=GEN all
lebu=ki=te *diyo=ga* *mu?u* *mu?u=te* *no?=na* *mã?e=yo?*
 person=PL=OBL daily=FOC one.HUM REP=OBL eat=INF begin=ACT.PST
 ‘It is believed (= [*is*] *the matter of the mind*) that one day, a great demon suddenly appeared from somewhere and began to eat all the people of the village, one by one.’

Here again, the speaker is presenting a general, unquestionable truth, hence there is no element translating as ‘be’ What is being narrated here is not THAT people believe something but rather WHAT they believe.

Summarizing, as the preceding pages have shown, there is a clear semantic difference between qualitative predication with and without an element translating as ‘be’ in English. In other words, the absence of *heke* or *ayij* in Kharia is semantically relevant and cases of qualitative predication in which there is no element translating as ‘be’ cannot be viewed as having a zero copula.

Structurally speaking, qualitative predication with an element translating as ‘be’ is similar to dynamic narrative predication discussed in 6.1: In both types of predication we find an element marked for TAM and person translating into English as a verb. Also, as noted above, qualitative predication with an element translating as ‘be’ is more “narrative” in style than qualitative predication without this element. The only difference between the two “narrative” types of predicates, structurally speaking, is that with “narrative” qualitative predication, the status of the present tense is somewhat different (“here and now”) as compared with the more general present tense of dynamic narrative predication. Also, the present-tense forms in qualitative predication do not mark for basic voice, whereas those for dynamic narrative predication obligatorily mark for basic voice. Otherwise, the two types of predication are structurally quite similar, in opposition to the (non-narrative) type of qualitative predication in which there is no element marked for TAM, basic voice or person.

7.4 CLAUSE COORDINATION

The coordination of two or more finite clauses in Kharia is achieved through any of a large number of conjunctions. Their presence has no effect on word order. The following list presents a few of the most common forms, followed by two brief examples.

107. *ro, oqo(?)* ‘and’ *la?* ‘(and) then’
muda, lekin, magari ‘but’ *la?ko* ‘but’
umla?, umla?ko, umbo? la?ko ‘otherwise’
108. *ho konthe²d=te yo=yo?* *ro botoŋ go²d=ki* *ho=kaŋ.*
 that bird=OBL see=ACT.PST and fear C:TEL=Mid.PST that=SG.HUM
 ‘He saw that bird and became afraid.’ [AK, 1:41]

109. *musniṅ dimu bheir lam=oʔ=ki muda umay kuy=oʔ.*
 once day entire search=ACT.PST=PL but NEG.3PL find=ACT.PST
 ‘Once [they] searched all day but didn’t find [the demon].’
 [MT, 1:68]

7.5 SUBORDINATION

This section deals with various types of subordination in Kharia, including “adverbial clauses” in 7.5.1 (purpose, causal / instrumental, temporal, conditional and counterfactual, concessive and depictive) and complement clauses in 7.5.2. As there are a large number of productive strategies for expressing propositional attribution (“relative clauses”), these will be dealt with separately in 7.6.

7.5.1 “Adverbial” clauses

7.5.1.1 Purpose clauses

Purpose clauses are constructed with the infinitive and usually one of the following two postpositions: *ghaʔd* ‘for; PURP’ and *thon* ‘for; PURP’. With these postpositions, the infinitive is usually marked for the genitive.

110. *u khoṛi boʔ=ki=te lebu=ki ek dusre=te sāghro*
 this village place=PL=OBL person=PL one second=OBL help
karay=na=? ghaḍ, apan sistam oʔj=siʔ=may.
 do=INF=GEN PURP REFL system take.out=PERF=3PL
 ‘The people of these villages, *in order to help one another*, have developed (= taken out) their own system.’ [AK, 5:11]
111. *laʔ souʔb=ga bhai=ki ho khajar tar=na ghaḍ juda*
 then all=FOC brother=PL that deer kill=INF PURP separate
juda muʔ=ki=may, kinir=te.
 REP emerge=MD.PST=3PL forest=OBL
 ‘Then all the brothers set out separately *to kill that deer*, into the forest.’ [AK, 1:15]

Occasionally, the infinitive alone is used, especially with contentive morphemes denoting motion:

112. ... *ber=a?* *ore'j ayi'j* *harhowa=ki a'j=may ho=ki*
 who=GEN OX QUAL.PRS plower=PL QUAL.PRS=3PL that=PL
apan har *do'd=ke silo?=na col=ta=may.*
 REFL plowing.utensils take=SEQ plow=INF go=Mid.PRS=3PL
 someone has oxen, [and] there are plowers, they take their own
 plowing utensils and *go to plow.* [AK, 5:15]

7.5.1.2 Causal / instrumental clauses

The most common means of expressing a causal clause is through the subordinating conjunction *ina no* [why CMPL] 'because'. In these clauses, the TAM/PERSON-syntagma is always morphologically fully finite.

113. *kundqab aw=ki tomliŋ khariya gam dom=na*
 behind QUAL=Mid.PST milk Kharia say PASS=INF
la?=ki=may ina no u=ki tomliŋ u'd=ga del=ki=may.
 IPFV=Mid.PST=3PL because this=PL milk drink=FOC come=Mid.PST=3PL
 'Those who were behind were called "Milk Kharia" *because they came drinking milk.* [MT, 1:180]

Another common means of denoting causal or instrumental clauses is through the infinitive, followed by an ablative postposition:

114. *mehneit ro iman=ya? kamu karay=na se idi'b=te*
 labor and honor=GEN work do=INF ABL night=OBL
su?kho=ya? leme'd del=ta.
 happiness=GEN sleep come=Mid.PRS
 'By doing labor and honourable work (= from labor and honor's working), one sleeps happily (= the sleep of happiness comes) at night.' [BB, 2:67]

7.5.1.3 Temporal clauses

To express that one event took place before or after another event, either the postposition *sej* (or its alternative form *siŋ*) 'before' or *lo'dqho* or *badte*, both with the meaning 'after', is used. The predicate of the temporal clause appears in the infinitive, generally in the genitive case.

115. *co=na=? siŋ burha aqi=ya? be't=qom=ki=te u*
 go=INF=GEN before old.man ANAPH=GEN son=3POSS=PL=OBL this

gam=oʔ:

say=ACT.PST

‘Before [their] going, the elderly man said this to his sons:

[MT, 1:130]

116. *ud=na loʔdho ađi ho maha daru tuta=te=ga*
 drink=INF after ANAPH that big tree bottom=OBL=FOC

đel=ki.

come=MID.PST

‘After drinking [the water], he came to the bottom of that big tree.’

[AK, 1:23]

To express that one event takes place as soon as another takes place, the predicate appears as an infinitive and is directly followed by *sariyat* / *seriyat*, a loan word from Sadri:

117. *ondor=na sariyat beđa, beđi, konselduʔ oʔ điyar*
 hear=INF as.soon.as boy girl woman house enter

goʔđ=ki=may...

C:TEL=MID.PST=3PL

‘As soon as [they] heard [this], the children and women entered [their] homes.’

[Kerkeṭṭā, 1990: 2]

In another construction with the same meaning, the conjunction *noga* ‘as soon as’ follows the subordinate clause, which contains a fully finite TAM/PERSON-syntagma. *noga* apparently derives from the complementizer *no* and the focal marker *ga*.

118. *ro sađa ondoro=te noga jor jor se*
 and sound hear=ACT.PRS as.soon.as strong REP INST

bhabru=te.

bark=ACT.PRS

‘And as soon as he hears a sound, [the dog] barks loudly.’

[BB, 3:37]

Alternatively, the infinitive is followed by *deri* to denote ‘as soon as’ or ‘while’:

119. oʔ dɣyar=na deri bij=dɔm=te jun=oʔ
 house enter=INF while daughter=3POSS=OBL ask=ACT.PST
 “gach=ob no beʃi?”
 fry=ACT.PST.2SG Q daughter
 ‘While entering the house, she asked her daughter “Did you fry him,
 daughter?”’ [BB, 1:79]

120. gitaʔ=na deri poroŋ=te ikud jughay lemed
 lie.down=INF as.soon.as hare=OBL very much sleep
 laʔ=ki ro lemed goʔd=ki.
 EMOT=MID.PST and sleep C:TEL=MID.PST
 ‘As soon as he lay down, the hare became very tired and fell asleep.’
 [BB, 3:16]

Similarly, the correlative conjunction *caʃ... paʃ* may be used, with the same meaning:

121. modi=yaʔ caʃ mǎgni paʃ biha
 Modi=GEN as.soon.as water.ceremony as.soon.as marry
 hoy=ki.
 become=MID.PST
 ‘No sooner had Modi’s water ceremony taken place did the wed-
 ding take place.’ [RD, 2:47]

Another very common means of denoting a temporal clause is through a “relative” clause of the type discussed in 7.6 in which the “head” is a generic time word, generally *bhere* ‘time’, which is modified by a preposed clause with a finite TAM/PERSON-syntagma.

122. modi cautha klas=te aw=ki bhere ap=dɔm moŋ
 Modi fourth class=OBL QUAL=MID.PST time father=3POSS one
 saykil soŋ kay=oʔ.
 bicycle buy BEN=ACT.PST
 ‘When Modi was in the fourth class, his father bought him a bicycle.’
 [RD, 2:17]

123. modi paguʃ rēgse=na laʔ=ki bhere tay=ga arkhi
 Modi crawl crawl=INF IPFV=MID.PST time ABL=FOC liquor

u²d=na=? thoŋ ñyam=na la²=ki.
 drink=INF=GEN PURP cry=INF IPFV=MID.PST

‘Since the time that Modi was crawling he used to cry because he wanted to drink liquor (= cry in order to drink liquor).’

[RD, 2:5]

To express ‘until’, *jaw-* ‘until’ forms a compound with the lexical base of the predicate. The subject of the corresponding finite clause can appear here in the genitive, all other arguments / adjuncts have the same form as in a finite clause.

124. *anita=? jaw-dam, . . .* ‘Until Anita comes, or ‘Until Anita’s
 Anita=GEN until-arrive arrival, .

For further strategies for expressing temporal clauses in Kharia, see the discussion of sequential and imperfective converbs in Sections 6.6.2.3–6.6.2.4.

7.5.1.4 Conditionals and counterfactuals

In both conditional and counterfactual clauses, the apodosis is almost always marked by the clause-initial conjunction *la²* ‘then’. The Indo-Aryan loan word *agar* ‘if’ can appear in the protasis, although it is not generally used. Very rarely, *yadi* ‘if’, a borrowing from modern Indo-Aryan (where it is a borrowing from Sanskrit), is found.

The TAM/PERSON-syntagma of both the protasis and apodosis generally appears in the irrealis, less commonly in the irrealis perfect, especially in counterfactuals, which are otherwise structurally identical.

Conditionals

125. *am am=a² ti² buŋ ob-dhog=e=m la² sebol*
 2SG 2SG=GEN hand INST CAUS-grab=ACT.IRR=2SG then delicious

la²=na.

EMOT=MID.IRR

‘If you give me it (= cause [me] to grab [it]) with your [own] hands, then [it] will taste delicious.’ [BB, 1:43]

126. *lemed um la²=ta la² korob korob=ga*
 sleep NEG EMOT=MID.PRS then silent REP=FOC

gitaʔ=siʔ=na.

lie.DOWN=PERF=MID.IRR

‘If you are not tired then lie quietly.’

[BB, 3:22]

127. *agar am=aʔ bhabru onqor=kon kiʃoʔ de=na laʔ*
if 2SG=GEN bark hear=SEQ tiger come=MID.IRR then

boriya=te=ga tar nog=e.

both=OBL=FOC kill eat=ACT.IRR

‘If, having heard your barking, the tiger comes, he will kill and eat [us] both.’

[BB, 3:40]

Occasionally, the conjunction *tob(=ga)* ‘then’ from Indo-Aryan is found in place of the more common *laʔ* ‘then’ in the apodosis.

128. *muda in=te moʃ hapta=te am moʃ hajar rupaya*
but 1SG=OBL one week=OBL 2SG one thousand rupees

kamu=ke ter=e=m

earn=SEQ give=ACT.IRR=2SG

‘But if you earn and give me 1,000 rupees in one week,’

tob=ga in kãdij=na=in no am hoqom kamu
then=FOC 1SG accept=MID.IRR=1SG Cmpl 2SG other work

karay=kon borol=na=m.

do=SEQ live=MID.IRR=2SG

‘only then will I accept that you will live by doing other work.’

[BB, 2:51]

Although quite rare, the corpus also contains examples such as the following, both from the same person, where the TAM/PERSON-syntagma in the apodosis appears in the past imperfective instead of the irrealis to denote a habitual event.

129. *ap=qom=te botol dhoʔ-dhoʔ yo=ye laʔ*
father=3POSS=OBL bottle take-RDP see=ACT.IRR then

cokoʔd=ga guʃiyay khor=na laʔ=ki.

drag.oneself=FOC crawl ITER=INF IPFV=MID.PST

‘If he sees his father take the bottle then he *would crawl around here and there.*’

[RD, 2:13]

130. *oʔ=te ap=dom teinko arkhi un=sig=e*
 house=OBL father=3POSS a.little liquor place=PERF=ACT.IRR
ho=jeʔ=ko ho=kaʔ capu karbaʔd=kon uʔd
 that=SG.NHUM=CNTR that=SG.HUM rummage ECHO=SEQ drink
idaʔ=na laʔ=ki
 ECHO=INF IPFV=MID.PST
 ‘If his father has placed some liquor in the house, then he *would*
 rummage through everything [until he found it] and then *drink it all*
up.’ [RD, 2:25]

Counterfactuals

131. *agar in hante aw=siʔ=na=in laʔ ho=kaʔ=aʔ*
 if 1SG there QUAL=PERF=MID.IRR=1SG then that=SG.HUM=GEN
kayom onʔor[=e]=in.
 speech hear[=ACT.IRR]=1SG
 ‘If I had been there I would have listened to His (= God’s) words.’
 [HJPa:103]

7.5.1.5 *Concessives*

The concessive construction in Kharia is not structurally related to the conditional construction. Concessive clauses are formed by means of the correlative construction *jawbhi tawbhi* ‘although. . . nevertheless’, which has been borrowed from Indo-Aryan. *tewjo* or *tewaʔjo* ‘nevertheless’¹⁷ is also found instead of *tawbhi*. The predicate of the subordinate clause is fully finite.

132. *jawbhi ho=kaʔ dhangar ayij, tawbhi*
 although that=SG.HUM servant QUAL.PRS still
ho=kaʔ=aʔ=te laij besu peʔ um kui=taj.
 that=SG.HUM=GEN=OBL stomach full rice NEG find=MID.PROG
 ‘Although he is a servant, still he goes hungry (= a stomach-full of
 rice isn’t being found to him).’ [HJPa:60]

¹⁷ From Sadri *teu* ‘but’ plus the Kharia focal marker =*aʔ* and the additive focal marker =*jo*.

7.5.1.6 *Depictives*

Following Schultze-Berndt & Himmelmann (2004: 77f.), depictives, or “depictive secondary predicates”, may be defined as constructions exhibiting the following traits:

- The construction contains two separate predicating elements—the finite predicate and the depictive—whereby the eventuality of the depictive predicate holds for the time-frame of the eventuality of the finite predicate;
- The depictive predicate is obligatorily controlled, although this controller may not be overtly expressed as an argument of the depictive;
- The predication of this controller by the depictive is at least partially independent of that of the finite predicate—i.e., the two together do not form a complex predicate;
- The depictive is not an argument of the finite predicate, i.e., it is never obligatory;
- The depictive is not a modifier of the controller (e.g., it is not an adjective);
- The depictive is non-finite. Although this is certainly a problematic criterion from a cross-linguistic perspective (cf. e.g., Koptjevskaja-Tamm, 1994; Maas, 2004), the criterion for this trait given by Schultze-Berndt & Himmelmann, namely that the depictive predicate is not marked for tense or mood categories, will suffice for our purposes;
- The depictive is part of the same prosodic unit as the finite predicate.

There are a number of constructions in Kharia which can express the same meaning as depictives defined in this sense. Most of these are not “genuine” depictive constructions in that they are also productively used to express other types of “adverbials” One possibility is to use the “sequential” converb, discussed in Section 6.6.2.3:

133. *hin tay=ko* *soub=ga mon oʔ=te* *aw=kon*
 that ABL(= ‘since then’)=CNTR all=FOC one house=OBL live=SEQ
- sori sori kamu=kon borol=na laʔ=ki=may.*
 together REP work=SEQ live=INF IPFV= MID.PST=3PL
- ‘Since then all of them lived in one house, working together for their living [= *having lived and having worked together*].’

[RD, 2:114]

Although this converb's primary function is to denote an action which begins before the eventuality expressed by the finite TAM/PERSON-syntagma, examples such as (133), in which the two eventualities are simultaneous, are not uncommon and are semantic depictives. However, as their primary function is to express sequential events, they cannot be considered a genuine depictive construction. In addition, sequential converbs need not have the same subject as the main predicate, another requirement of depictives.

Another, more common means of expressing semantic depictives in Kharia are the translational equivalents of imperfective converbs (6.6.2.4), which often have a depictive interpretation:

134. *muda ekle* aw=ga awga *muruk ansa* laʔ=ki.
 but alone QUAL=FOC REP very unhappiness EMOT=MID.PST
 'But [He (= God)] became unhappy being alone (= *being* alone, He became unhappy).' [AK, 3:6]

Although such constructions always seem to have the same subject as the finite TAM/PERSON-syntagma, they are quite often separated from the main predicate and then form an independent phonological unit.

Another means of expressing semantic depictives is through a masdar marked for the oblique case:

135. *khõqeʔj=ko* ɖoko=te=ga *mukum=ki,* *tay ho boʔ=te=ga*
 a.while=CNTR sit=OBL=FOC doze.off=MID.PST then that place=OBL=FOC
 "*khõqeʔj mene ɖhalanʃe=ta=ɨn*" – *gam=oʔ ro*
 a.while CLASS take.a.nap=MID.PRS=1SG say=ACT.PST and
gitaʔ=kon lemeʔɖ god=ki.
 lie.down=SEQ sleep C:TEL=MID.PST
 'Sitting for a while, he began to doze off (= he dozed off), so he said [to himself] at that place "I'll sleep for a while." and he lay down and fell asleep.' [BB, 2:27]

Again, this is not a genuine depictive construction, as this construction may have a different "subject" (= S/A) than the finite TAM/PERSON-syntagma which may also be overtly expressed (in the genitive). It is also found primarily in attributive function.

Thus the last three constructions, although often compatible with a depictive interpretation, must be considered general adjunct constructions.

The case is somewhat different with then next construction, whose primary and perhaps only function is to express semantic depictives: The eventuality of the non-finite predicate is temporally co-extensive with that of the finite TAM/PERSON-syntagma and is restricted to the same subject as that of the finite TAM/PERSON-syntagma.

This construction consists of what at first glance appears to be a predicate with a complex semantic base (6.3.2), but in which the first predicating sub-base(s) is/are separated from the final base by a marker which is homophonous with the focal particle =*ga*. The finite TAM/PERSON-syntagma usually contains a semantic base denoting motion.

136. *ho=ki* *lere?* *laḏa* *jayḏam=ga* *ḏel=ki=may.* *ho=ki*
 that=PL rejoice laugh be.eager=FOC come=MID.PST=3PL that=PL
aṛmaṛay=ga *siḏ=ga* *ḏel=ki=may.*
 hesitate=FOC get.lost=FOC come=MID.PST=3PL
 ‘They came *rejoicing, laughing and eager*. They came *hesitating and searching* (i.e., some came eagerly, some hesitantly).’
 [MT, 1:4,5]

137. *kunḏab* *aw=ki* *tomliṅ* *khariya* *gam* *ḏom=na*
 behind QUAL=MID.PST milk Kharia say PASS=INF
la?²=ki=may *ina no* *u=ki* *tomliṅ* *u’ḏ=ga* *ḏel=ki=may.*
 IPFV=MID.PST=3PL because this=PL milk drink=FOC come=MID.PST=3PL
 ‘Those in the rear were called “Milk Kharia” because they came *drinking milk*.’
 [MT, 1:180]

In the following example, the object of the depictive predicate, a complement clause, is postposed, as it is a “heavy” constituent.

138. ... *u* *goḏjun=te* *socay=ga* *col=ki=ki* *no* “*i* *jāut*
 this path=OBL think=FOC go=MID.PST=PL Cmpl what animal
heke *hoy?*”
 QUAL.PRS INFER
 ‘ they walked along this path, *thinking* “*What animal could it be?*” ’
 [AK, 1:12]

In the following example, the depictive predicate again appears before the finite predicating element, thereby separating that element from its goal:

139. *tay, ikon, ho=ki apan moʔjhi=te kayom=ki=may no*
 then umh that=PL REFL middle=OBL speak=MID.PST=3PL CMPL

“*a=kaʃ seŋ ɖaʔ kuy=e*
 CREL=SG.HUM first water find=ACT.IRR

ho=kaʃ u daru=te=ga yo=ga ɖe=naʔ.
 that=SG.HUM this tree=OBL=FOC see=FOC come=MID.IRR

‘Then, umh, they said amongst themselves “He who first finds water, he should come to this tree, *looking* [for the others].” ’

[AK, 1:20]

However, as the following examples show, the depictive co-predicate is not restricted to the position directly preceding the finite TAM/PERSON-syntagma. This is problematic for considering this a genuine depictive construction, at least in the definition used here, as such depictive predicates are not within the same prosodic unit as the main predicate:

140. *khariya dhirom dhirom buli=ga aʃmaʃay=ga kiʃoʔ,*
 Kharia slowly REP wander=FOC hesitate=FOC tiger

buwan, bichi buŋ kheʔd ɖuŋ=ga jol ɖuŋ=ga
 snake scorpion INST bite PASS=FOC ECHO PASS=FOC

ɖel=ki=may.

come=MID.PST=3PL

‘The Kharia came, *wandering* slowly, *stopping*, *being bitten*, etc., by tigers, snakes and scorpions.’

[MT, 1:13]

There is a very close relationship—both morphosyntactically and semantically—between one of the constructions discussed in Section 6.6.2.4 which function as “imperfective converbs”, i.e., the repetition of a lexical morpheme marked by =*ga*, as in (134), and the “depictive” construction described here. Morphosyntactically, the two constructions have in common that they both are (or can be) marked by the focus marker =*ga*, the difference between the two being that the “imperfective converb” is repeated, whereas the depictive construction is not.

Repetition of the lexical morpheme in the imperfective converb is clearly an iconic means of signaling iterativity or extended duration and its use would also seem to emphasize the fact that two more-or-less independent eventualities simultaneously hold. The non-reduplicated

depictive construction, on the other hand, closely resembles structurally a TAM/PERSON-syntagma with a complex semantic base.

The status of =ga in the depictive construction is somewhat unclear: As the glosses in the examples above show, =ga is homophonous with the focus marker =ga (7.7.4). This would seem to make sense, as “depictives” (as well as “imperfective converbs”) virtually always focus on the information of the secondary predicate (which is the “primary” predicate from a pragmatic point of view). It would be wrong, however, to consider it merely a focus marker in the depictive construction: If this unit is not repeated and is not marked for =ga, we have a complex semantic base, consisting of two sub-bases. Although not a hard and fast rule, predicates of this type are generally interpreted as being sequential, and when asked, informants agreed that the presence or absence of =ga yields very different interpretations:

141. *tomliŋ u²d de=na!* ‘Drink milk [and then] come!’
 milk drink come=MID.IRR’,

142. *tomliŋ u²d=ga de=na!* ‘Come drinking milk!’
 milk drink=FOC come=MID.IRR

As the conditions which combine to determine whether a TAM/PERSON-syntagma whose semantic base consists of two (or more) sub-bases is to be interpreted as sequential or simultaneous are not entirely clear, no further comments can be made at this point on the status of =ga in this construction, other than to note the close connection between focus, imperfective converbs and depictives.

The depictive construction is negated by *enem* ‘without’ =ga is then optional.

143. *ho=kaŋ=a² pura kayom enem on²or=ga gam=o² no:...*
 that=SG.HUM=GEN all word without hear=FOC say=ACT.PST CMPL
 ‘But *without hearing* all his words he said. [HJPa:68]

This also holds in those cases where the “depictive” predicate is not directly adjacent to the finite TAM/PERSON-syntagma. Thus, the same strategy is used to negate both imperfective converbs and depictives:

144. *iđib bhere enem bhabru solo? um aw=na pal=na*
 night time without bark dog NEG remain=INF be.able=INF

la?=ki.

IPFV=MID.PST

‘At night the dog could not keep quiet (= could not remain *without barking*).’ [BB, 3:19]

Finally, as the last example and the following examples show, =*ga* is not obligatory in the negated depictive construction, unlike in the non-negated form:

145. *jab ho=ki ciŕra kui=na la?=ki=may la? enem*
 CREL.TEMP that=PL squirrel find=INF IPFV=MID.PST=PL then without

tar *umay melay=na la?=ki.*

kill NEG.3PL leave=INF IPFV=MID.PST

‘Whenever they found a squirrel, then *without killing [it]*, they wouldn’t leave.’ [HJpa:43]

146. *musa=ko in enem daichna ɖoʔɖ um=in melay[=e].*
 today=CNTR 1SG without fee take NEG=1SG leave=ACT.IRR

‘Today, *without receiving my fee*, I will not leave.’

[Kerkeṭṭā, 1990: 3]

In sum, the construction referred to here as the depictive construction differs from both complex predication as well as non-negated imperfective converbs, although only minimally from both. This fits in well with the fact that it is, semantically speaking, closely related to both.

Predicates of motion—path and manner Kharia appears to belong to what Talmy (1985) refers to as “Spanish-type” languages (including Romance in general, Semitic, Polynesian, Nez Perce, Caddo, but also apparently Hindi, cf. Narasimhan, 2003) in that lexical items denote motion and path but, if manner is to be expressed, it must be expressed through the depictive construction.

The following lexical items combine the notions of motion and path. This list is probably not exhaustive, although there are presumably not many more contentive morphemes in this category.

147. *mu?* ‘emerge’ (motion out of an object)
qiyar ‘enter’ (motion into an object)
col ‘go’ (motion away from deictic center (usually the speaker))
qel ‘come’ (motion towards deictic center)
qam ‘arrive’ (motion arriving at the deictic center)

As we shall see, these morphemes may also be combined with (telic) goals. The following may also perhaps belong to this category:

148. *qeʔb* ‘rise, ascend’
eŋ ‘return’
tuta ‘lower’ (ITR/MIDDLE; TR/ACTIVE)
tobhlun ‘rise’ (MIDDLE); ‘raise’ (ACTIVE)

On the other hand, when other contentive morphemes which translate in English as verbs of motion denoting manner are combined with a (telic) goal, these must be expressed through the depictive construction, often with the reduplication of the depictive predicate.

149. **sahar=te*¹⁸ / **khoʔtay dhãy=o?*. ‘S/he ran to / up to the city.’
city=OBL up.to run=ACT.PST

150. *sahar bhiʔtarte dhãy=ga (dhãy=ga) qiyar=ki.*
city within run=FOC REP enter=MID.PST
‘S/he entered the city, running.’

151. **khoʔi=te*¹⁹ / **khoʔtay buli=ki.*
village.section=OBL up.to wander=MID.PST
‘S/he wandered to / up to the village.’

152. *buli buli=ga sahar=te qiyar=ki.*
wander REP=FOC city=OBL enter=MID.PST
‘S/he wandered into the city / entered the city, wandering.’

¹⁸ This sentence is grammatical, however, with the meaning ‘S/he ran in the city.’ See discussion below.

¹⁹ See note 18.

153. **laphŋga bahar tay laphŋga=te paure=ki.*
 cave outside ABL cave=OBL swim=MID.PST
 ‘S/he swam from outside the cave into the cave.’

154. *paure paure=ga laphŋga=te dīyar=ki.*
 swim swim=FOC cave=OBL enter=MID.PST
 ‘S/he swam into the cave / entered the cave, swimming.’

However, if what is being described is motion AWAY from something or motion WITHIN something, i.e., an atelic path, then manner may be expressed together with motion, and the path is expressed by means of a postposition or case marker.

155. *sahar=te / tay dhāy=o?* ‘S/he ran in / from the city.’
 city=OBL ABL run=ACT.PST

156. *laphŋga tay paure=ki.* ‘S/he swam out of the cave.’
 cave ABL swim=MID.PST

7.5.2 Complement clauses

7.5.2.1 In subject function

Complement clauses in subject function are rare.²⁰ The predicate is often an infinitive, as in (157).

157. *oreʔj koŋtaŋ bui=na=ko ho=ki=yaʔ dhatam aw=ki.*
 ox cow raise=INF=CNTR that=PL=GEN custom QUAL=MID.PST
 ‘But raising oxen and cows was their custom.’ [MT, 1:9]

There are also examples such as the following two, with a finite TAM/PERSON-syntagma, in which the entire complement clause is marked by the postposition *guḍ* ‘like’ (158), although this is not necessary (159).

158. *menson=ko um ḍaŋ=e=m guḍ laʔ=ta.*
 once=CNTR NEG send=ACT.IRR=2SG like EMOT=MID.PRS
 ‘It looks like you won’t repel (= send) them all at once.’
 [Kerkeṭṭā, 1990: 11]

²⁰ Or rather, complement clauses which refer to the enclitic subject marker, in this case “zero” (3sg), are rare.

159. am um co=na=m la[?]=ta.
 2SG NEG go=MID.IRR=2SG EMOT=MID.PRS
 ‘It looks like *you won’t be going.*’

7.5.2.2 In object function

Complement clauses in object function are quite common, especially with predicates of speech. These clauses have a finite TAM/PERSON-syntagma and usually begin with the complementizer *no*. Speech and thought are generally presented as they would have been uttered by the person who said or thought them. *no* can be omitted.

160. ap=d_{om} ra_{ta}=te remakh=o[?] ro gam=o[?] no “babu
 father=3POSS Rata=OBL call=ACT.PST and say=ACT.PST Cmpl child(voc)
 musa i_n kinir co=na um=i_n pal=e.”
 today 1sg forest go=INF NEG=1sg be.able=ACT.IRR
 ‘His father called Rata and said [to him] “*Son, today I will not be able to go to the forest.*” ’ [BB, 2:21]

Instead of *no*, *gam=k_{on}*, the sequential converb of *gam* ‘say’, is occasionally found clause-finally as a kind of quotative particle:

161. ho=ghay=ga ho di_{nu}=jo rel=ya[?] le²bq_{om} pōga bajay=o[?]
 that=way=FOC that day=ADD train=GEN boss horn sound=ACT.PST
 la[?]=ko “malik=ga pōga bajay=o[?]” gam=k_{on} soub
 then=CNTR(= ‘but’) boss=FOC horn sound=ACT.PST say=SEQ all
 rel pa_{tari}=te col go²d=ki=may.
 train track=OBL go C:TEL=MID.PST=3PL
 ‘In just that way, on that day as well the boss of the train sounded his horn, but *thinking* (= *having said*) “*The boss has sounded the horn.*”, all [the sheep and goats] went off to the train track.’ [RD, 1:13]

162. “je janwar tar=e=ki ho janwar=ya[?] ghos
 CREL animal kill=ACT.IRR=PL that animal=GEN meat
 ol=e=ki” gam=k_{on} gam=o[?], swapan, munu buj.
 take=ACT.IRR=PL say=SEQ say=ACT.PST dream dream INST
 ‘“*Whatever animal they kill, that animal’s meat they should bring [back to you]*” he said, in (= through) a dream.’ [MS, 2:8]

Occasionally, *gam=kon* is also found in other functions, although the data at the moment are rather scarce. It would appear to function as a focal particle in these constructions, although this is far from certain.

163. *ho=kaʔ=aʔ thoŋ moŋ toŋliŋ gam=kon dudharu*
 that=SG.HUM=GEN for one milk say=SEQ milk.bearing
gay=jo ho=ki melay tu=yoʔ=ki.
 COW=ADD that=PL leave DPT=ACT.PST=PL
 ‘For him [i.e., the small child] they also left a milk cow behind for milk (= one *having said* “milk” milk-cow).’ [MT, 1:159]
164. *moŋ beʔa gam=kon ho=kaʔ=te rakam rakam rāg ritaʔ*
 one son say=SEQ that=SG.HUM=OBL way REP color ECHO
lutui phaʔa ob-su=na laʔ=ki=kiyar.
 clothing ECHO CAUS-wear=INF IPFV=MID.PST=DU
 ‘Seeing as he was their son (= *having said* “a son”), they dressed him in all kinds of brightly colored clothes.’ [RD, 2:3]

Direct speech may also be rendered without a complementizer:

165. “am=aʔ thoŋ u karay=na . karay kay[=e]=iŋ”
 2SG=GEN for this do=INF do BEN=ACT.IRR=1SG
gam=e.
 say=ACT.IRR
 ‘He will say “I will do this for you” ’ [MS, 1:282]

Another contentive morpheme which commonly takes a complement clause is *koŋ* ‘know, find out’:

166. *muda tam jou lebu=ki umay koŋ=sikh=aʔ*
 but now up.to person=PL NEG.3PL KNOW=PERF=ACT.PST
 no u=jeʔ i daru heke.
 Cmpl this=SG.NHUM what tree QUAL.PRS
 ‘But to this day, people have not found out *what tree it is.*’
 [BB, 1:18]

Finally, complement clauses can be expressed with a masdar as the predicate, e.g., when the matrix predicate denotes perception, as in the following two examples.

167. ap=ḍom=te botol dhoʔ-dhoʔ yo=ye laʔ
 father=3POSS=OBL bottle take-RDP see=ACT.IRR then
- cokoʔq=ga guʔiyay khor=na laʔ=ki.*
 drag.oneself=FOC crawl ITER=INF IPFV=MID.PST
- ‘If he sees *his father take the bottle* then he would crawl around here and there.’ [RD, 2:13]

168. modi=yaʔ uq̣-uq̣ yo=kon ap=ḍom konoʔq=na
 Modi=GEN drink-RDP see=SEQ father=3POSS think=INF
- laʔ=ki no “beʔa=iʃn i=ghay buʃ roʔ sudhre=na?”*
 IPFV=MID.PST Cmpl son=1SG what=way INST FOC improve=MID.IRR
- ‘Seeing *Modi’s drinking*, his father would think “How will my son improve?”’ [RD, 2:38]

7.6 “RELATIVE CLAUSES”—PROPOSITIONAL ATTRIBUTION

In the present section, a detailed discussion of propositional attribution in Kharia is presented, constructions which correspond to relative clauses in languages such as English. However, the use of the term “relative clauses” will generally be avoided here, as this is less accurate for a number of reasons which will become apparent in the following pages.

Perhaps the main argument against referring to propositional attribution in Kharia simply as relative clauses is that only some of the constructions discussed in the following pages may be considered internal to a Case-syntagma, whereas others, most notably the correlative constructions and “circumnominal” clauses, are external to a Case-syntagma.

No systematic study has been made here with respect to the differences and similarities between restrictive and non-restrictive attributive clauses in Kharia: As almost all of the examples in the corpus are of restrictive clauses, there is simply too little data on non-restrictive or appositive clauses to make any general statements about them. The following presents a few examples for non-restrictive clauses, both elicited and from the corpus, all with a modified proform.

169. lebu=ki=yaʔ yo-yo *ɪn* *peʔ* *noʔkh=oʔj*.
 person=PL=GEN see-RDP 1sg rice eat=ACT.PST.1sg
 ‘I, *who the people saw*, ate rice.’
170. lebu=ki *ɪn=te* yo=yoʔ=ki *peʔ* *noʔkh=oʔj*.
 person=PL 1SG=OBL see=ACT.PST=PL rice eat=ACT.PST.1sg
 ‘I, *who the people saw*, ate rice.’
171. *oqo* *meson* *qe=na=?* *ray=na* *ro* *ray=kon*
 again once come=MID.IRR=FOC choose=INF and choose=SEQ
- ele=te* *je* *jhunq* *ayiʔj=le* *sori* *sori*
 1PL.EXCL=OBL CREL flock QUAL.PRS=1PL.EXCL together REP
- ele=te* *aqi=yaʔ* *boʔ=te* *qoq=e*.
 1PL.EXCL=OBL ANAPH=GEN place=OBL take=ACT.IRR
- ‘He [i.e., God] will come again to save us and, after saving us, *who are His flock*, He will take us all together to His home.’
- [MS, 1:314]

The following discussion is devoted almost entirely to restrictive constructions.

Strategies for propositional attribution

Of the four major strategies listed for attributive constructions in Keenan (1985), i.e. correlatives, prenominals, postnominals and internally headed attributive constructions, Kharia makes use of the following. In order to simplify terminology, we will speak here of prenominal, postnominal and circumnominal to avoid the somewhat cumbersome terminology “preposed to the lexical head of a Case-syntagma”, etc. This is only for practical reasons. These will be dealt with individually in the following pages.

- two different correlative constructions;
- nine different prenominal constructions;
- one circumnominal construction;
- the postnominal use of slightly modified forms of the two correlative constructions.

7.6.1 *Correlatives*

There are two different correlative constructions in Kharia:

- The “*je*-class” In this construction the correlative markers all begin with /j/ and have been borrowed from Indo-Aryan. Cf. the general correlative marker *je*, borrowed from Sadri. Other forms belonging to this class include forms such as *jab* ‘when (CREL)’, *jetna* ‘how much (CREL)’, etc. (cf., e.g., Biligiri, 1965: 100, §§1–3).
- The “*a* / *i* or *q*-class” This construction makes use of correlative forms which are homophonous with interrogatives, which generally consist of either the proclitic interrogative marker *a* = ‘*q*’ followed by a contentive morpheme, such as *a=ti²j* [*q*=side] ‘where?’, or its free-form *ata* ‘*q*’. There is also a small class of interrogatives which are based on the morpheme *i* ‘what?’, such as *i=ghay* [what=way] ‘how’ or *i bhere* [what time] ‘when’ (cf. 5.7 and also Biligiri, 1965: 102). Other interrogative forms found here include *be(ha)r* ‘who?’. This construction has not been borrowed from Indo-Aryan, although it could represent an older calque of the Indo-Aryan correlative construction using purely language-internal means.

In both constructions, the correlative marker appears directly before the lexical head, if this is overtly mentioned, most often at the beginning of the attributive phrase. The head is then often repeated in the main clause, preceded by a demonstrative. Alternatively, the head is not repeated, in which case only the demonstrative is found.²¹ The following provides examples for both constructions:

“*je*-class”

172. ... je khajar tar=sikh=øʔ=may **ho=kaʔ=aʔ** koməŋ=ko
 CREL deer kill=PERF=ACT.PST=3PL that=SG.HUM=GEN meat=CNTR

²¹ I find the data in Malhotra (1982: 312) somewhat perplexing in this respect, as well as her observation “The relative clause structure is generally attached to the right of the antecedent noun relativised; hence, relative clauses in Kharia are postnominal, the relative pronouns following the co-referential noun.”

While Malhotra’s observation is in accordance with her data, very few of the data I collected fit this description, as the vast majority of my data are examples of prenominals or correlatives of the types described below, with very few postnominals.

nalage, ...

NEG.QUAL.PRS

it isn't the meat of the deer that they had killed.

(*which deer they had killed, his* meat it is not. ') [AK, 1:59]

173. ...*ađi* je bhere eŋ=ki... *se bhere ađi=ya?*
 ANAPH CREL time return=MD.PST that time ANAPH=GEN

poŋom=te soreŋ kui=ki.

bundle=OBL stone find=MD.PST

when he returned, he found a stone in the bundle.'

(*which time he returned, that time. .*) [AK, 1:73]

“*a/i-class*”

174. ... a=boʔ=te pujapaṭh karay=na aw=ki, *ho boʔ=te*
 Q=place=OBL sacrifice do=INF QUAL=MD.PST that place=OBL

đam=ke ho=ki ho đoli=te maṛay=oʔ=may.

arrive=SEQ that=PL that palanquin=OBL put.down=ACT.PST=3PL

. having arrived **at the place where the sacrifice was to be done,**

they put the palanquin down.' [AK, 2:16]

The following presents two examples of “headless relative clauses”

175. *tay, ikon, ho=ki apan moʔjhi=te kayom=ki=may no*
 then umh that=PL REFL middle=OBL speak=MD.PST=3PL Cmpl

“a=kaṛ seŋ đaʔ kuy=e, *ho=kaṛ u*
 Q=SG.HUM first water find=ACT.IRR that=SG.HUM this

daru=te=ga yo=ga đe=na”.

tree=OBL=FOC see=FOC come=MD.IRR

‘Then, uh, they said amongst themselves “*He who first finds water,* **he** should come to this tree, looking [for the others]”’ [AK, 1:20]

176. a=te ate bes bes ṭhãŋo kuy=oʔ=ki, *hinte hinte khaṛiya*
 Q=OBL REP good REP place find=ACT.PST=PL there REP Kharia

goʔjlo? kamu=na laʔ=ki=may.

ricefield work=INF IPFV=MD.PST=3PL

‘*Wherever they found good land (= places), there* the Kharia would work the rice fields.’ [MT, 1:8]

The distribution of these two constructions is given in Table 7.6. In this and the following tables, only those categories which are listed are grammatical: Forms in parentheses “()” are grammatical for at least one speaker but not all, otherwise, the construction was acceptable to all.²²

Table 7.6: The distribution of the two correlative constructions

<i>“je-class”</i>				
$S_{REL} = S_{MAIN}$	$O_{REL} = O_{MAIN}$	$A_{REL} = A_{MAIN}$	Instrument _{REL}	(Genitive _{REL})
($S_{REL} = O_{MAIN}$)	$O_{REL} = S_{MAIN}$	$A_{REL} = S_{MAIN}$	Locative _{REL}	Comparative _{MAIN}
$S_{REL} = A_{MAIN}$	$O_{REL} = A_{MAIN}$	$A_{REL} = O_{MAIN}$	Temporal _{REL}	
<i>“q-class”</i>				
$S_{REL} = S_{MAIN}$	($O_{REL} = O_{MAIN}$)	$A_{REL} = A_{MAIN}$	Instrument _{REL}	Genitive _{REL}
$S_{REL} = O_{MAIN}$	($O_{REL} = S_{MAIN}$)	$A_{REL} = S_{MAIN}$	Locative _{REL}	Comparative _{MAIN}
$S_{REL} = A_{MAIN}$	($O_{REL} = A_{MAIN}$)	$A_{REL} = O_{MAIN}$	Temporal _{REL}	

7.6.2 *Prenominal constructions*

Common to all prenominal constructions is the fact that a Case-syntagma is marked according to its function in the main clause. According to the form of the predicate in the attributive clause, we may divide prenominals into nine categories, which will be discussed in groups in this section.

Prenominals along the finiteness continuum

The first three strategies to be discussed here can be considered as representing three discrete levels along a continuum of morphological finiteness of the predicate of the attributive clause.

Ia. Prenominals with a fully finite predicate Here, the attributive clause has the form of an independent clause with the exception of the head, which follows this clause. A simple example of this strategy can be found in the expression for ‘last’ (TEMP), e.g., *col=ki memon* [go=Mid.PST year] ‘last year’, i.e., ‘the year which went’ The following presents a few more examples. As (178) shows, these clauses are usually external to a Case-syntagma, as they precede all other elements in the semantic base of

²² Unfortunately, it was not possible to obtain information on the following forms: Instrument_{MAIN}, Locative_{MAIN}, Temporal_{MAIN} and Comparative_{REL}. Nevertheless, this and the following tables should give the reader a good impression of the respective productivity of these different constructions.

a Case-syntagma, including the demonstrative, although the demonstrative can precede this unit (cf. 5.8).

177. *ip yo=yoʃj lebu=ki ip=aʔ hotel=te aw=ta=ki.*
 1SG see=ACT.PST.1SG person=PL 1SG=GEN hotel=OBL live=MID.PRS=PL
 ‘The people I saw live in my hotel.’

178. *ip gam=sikh=oʃj ho=ghay=ga am=pe col=ki=pe, ...*
 1SG say=PERF=ACT.PST.1SG that=way=FOC 2=2PL go=MID.PST=2PL
 you went **just as (= the way) I had told you** ’ [AK, 1:63]

179. *ho=kaʃʔ ʃhik biha=na laʔ=ki=may bhere*
 that=SG.HUM exactly marry=INF IPFV=MID.PST=3PL time
qam=ki.
 arrive=MID.PST
 ‘He arrived just as they were getting married (= **at the they-were-just-getting-married time**).’ [HJPa:41]

This construction is also used with a demonstrative head functioning as a proform such as *u=jeʔ* [this=SG.NHUM] ‘this one’, *ho=jeʔ* [that=SG.NHUM] or simply *jeʔ* [SG.NHUM]: *ʔangay=siʔ (ho=)jeʔ* [hang=PERF (that=)SG.NHUM] ‘the one which has hung’ (adapted from Biligiri, 1965: 78).

Ib. Prenominals with a partially finite predicate The predicate of the attributive clause in this construction is partially finite as it lacks person-marking. Some examples (for comparison, the fully finite predicate is included in parentheses):

180. *ip=te yo=yoʔ lebu=ki ulaʔ likha=yoʔ=ki.*
 1SG=OBL see=ACT.PST person=PL letter write=ACT.PST=PL
 (cf. *yo=yoʔ=ki* [see=ACT.PST=PL])
 ‘The people who saw me wrote a letter.’

The distribution of these two constructions—i.e., attributive clauses with either fully finite or partially finite predicates—is given in Table 7.7.²³

²³ I consider these two constructions together here as their respective predicates are identical in the third person, singular and cannot be distinguished here for many examples in our data. Further research is necessary.

Table 7.7: The distribution of pronominals with a fully or partially finite predicate

(S _{REL} = S _{MAIN})	(O _{REL} = O _{MAIN})	(A _{REL} = A _{MAIN})		
(S _{REL} = O _{MAIN})	(O _{REL} = S _{MAIN})	(A _{REL} = S _{MAIN})	(Locative _{REL})	Comparative _{MAIN}
S _{REL} = A _{MAIN}		(A _{REL} = O _{MAIN})	Temporal _{REL}	

II. Pronominals with a masdar In these attributive constructions, the predicate of the attributive clause is a masdar in which polysyllabic forms consisting of an underlying contentive morpheme and any derivational marking appear in their bare form, while monosyllabic contentive morphemes are reduplicated.

As discussed in 6.6.2.1, what corresponds to the subject of a finite clause appears here in the genitive. Other arguments / adjuncts appear in the same form they would have in a finite clause. The following presents a few examples of this highly productive construction.

181. *ɨn=aʔ yo-yo lebu=ki ɨn=aʔ hoʔel=te aw=ta=ki.*
 1SG=GEN see-RDP person=PL 1SG=GEN hotel=OBL live=MID.PRS=PL
 ‘The people I saw live in my hotel.’

182. *adɨ=yaʔ su-su lutuy=te bhai=qom=te*
 ANAPH=GEN put.ON-RDP clothes=OBL brother=3POSS=OBL
oʔb-su=yoʔ. . .
 CAUS-put.ON=ACT.PST
 ‘He dressed his brother in **the clothes he himself was wearing** [HJPa:52]

183. *am=pe leŋ-leŋ jãut tar ol=e=pe.*
 2=2PL fly-RDP animal kill bring=ACT.IRR=2PL
 ‘You kill and bring [back] birds (= **animals which fly**).’ [MT, 1:131]

184. *ho=kaʔ ho=tay moŋ peʔ cakhnaʔ ro jhaʔi*
 that=SG.HUM that=ABL one rice curry and all
lekhe=naʔ ɲoʔ=na oʔb-soŋ oʔ=te qel qam=ki.
 kind=GEN eat=INF CAUS-buy house=OBL come arrive=MID.PST
 ‘From there he arrived at a restaurant (= **a house where rice, curries and foods of all kinds are sold**).’ [HJPa:66]

As noted in 6.6.2.1, the masdars have no inherent orientation, whether in terms of A, S, O, instrumental, locative, temporal, or tense. Consider once again the following examples:

O_{REL}

185. *ɪɲ=a?* *yo-yo* *lebu*
 1SG=GEN see-RDP person
 ‘**the person** *I saw / see / will see*’

A_{REL}

186. *ɪɲ=te* *yo-yo* *lebu*
 1SG=OBL see-RDP person
 ‘**the person** *who saw / sees / will see me*’

Instrument_{REL}

187. *ɪɲ=a?* *dura=te* *ru?-ru?* *kujji*
 1SG=GEN door=OBL open-RDP key
 ‘**the key** *I opened / open / will open the door with*’

Locative_{REL}

188. *ɪɲ=a?* *aw-aw* *ho?*
 1SG=GEN live-RDP house
 ‘**the house** *I lived / live / will live in*’

Despite this, there are certain limitations on these forms in actions involving two or more human or animate participants where ambiguities could arise as to the identity of the denoted actant, e.g., with causatives: While e.g. the intransitive morpheme *ɖiyar* ‘enter’ involves only one animate actant, its (slightly irregular) causative form, *ɖi-ʔbh>ar* [enter-<CAUS>] ‘have [someone] enter’ involves two animate actants, the primary and intermediate agents. In cases such as these, the masdar of the causativized form can refer to the house entered, the time of entry, etc., but the only person it may refer to is the primary agent, i.e., the person who had someone else enter, not the person who actually entered the house. To denote the intermediate agent with a causativized form, the passive form

is used, i.e., *q̄i*<?bh>*ar q̄om lebu* [enter-<CAUS> PASS person] ‘the person who was forced / made to enter’

Due to the fact that this form is completely underspecified with respect to TAM, it is most commonly found in environments in which the event it refers to is either presupposed or pragmatically accommodated. Reference to an event which is considered by the speaker to be neither of these is preferentially expressed through a different strategy, e.g., with a finite predicate, described in the preceding pages. The distribution of the construction with the masdar in attributive function is given in Table 7.8.

Table 7.8: The distribution of prenominal attributes with a masdar

$S_{REL} = S_{MAIN}$	$O_{REL} = O_{MAIN}$	$A_{REL} = A_{MAIN}$	Instrument _{REL}	
$S_{REL} = O_{MAIN}$	$O_{REL} = S_{MAIN}$	$A_{REL} = S_{MAIN}$	Locative _{REL}	Comparative _{MAIN}
$S_{REL} = A_{MAIN}$	$(O_{REL} = A_{MAIN})$	$A_{REL} = O_{MAIN}$	Temporal _{REL}	

The propositional attribute whose predicate is a masdar is internal to the semantic head of a Case-syntagma, as it follows determiners such as the demonstrative.

189. <i>soŋ-soŋ</i>	<i>lebu</i>	<i>ho</i>	<i>soŋ-soŋ</i>	<i>lebu</i>
buy-RDP	person	that	buy-RDP	person
‘buying ²⁴ man, merchant’		‘that merchant’	[Biligiri, 1965: 77]	

III. Prenominals whose predicate has the *-NV-* infix We can also present a brief discussion here of the prenominal construction whose predicate contains the *-NV-* infix (4.3.3). What would be the subject of the corresponding finite clause appears here in the genitive, as with the more productive masdar described in the last section.

As there are only two examples for this construction in our data, nothing can be said at present about its possible distribution, with the obvious exception that it can only be used with morphemes which are compatible with the *-NV-* infix. The following presents one example.

190. <i>khaŋiya=ki</i>	<i>ber=aʔ</i>	<i>ji<ni>b</i>	<i>peʔ=jo</i>	<i>umay</i>	<i>noʔ=na</i>
Kharia=PL	who=GEN	touch-<-NV->	rice=ADD	NEG.3PL	eat=INF
<i>laʔ=ki, . . .</i>					
IPFV=MID.PST					

²⁴ Biligiri here inadvertently translates *soŋ* as ‘sell’

‘The Kharia did not use to eat **rice** *which had been touched by anyone*,
[HJPa:122, ln. 57]

IV Infinitival Prenominals There are two prenominal constructions in which the predicate of the propositional attribute appears as an infinitive, which we will now deal with. They differ from each other formally only in terms of the case marking of the infinitive.

Prenominals with a genitive-marked infinitive

In this construction, the predicate of the propositional attribute is an infinitive which appears in the genitive.

191. *ro* [...] haḍa=na=ya? *bahan=a?* *ḍoli=te*
and pee=INF=GEN pretence=FOC palanquin=OBL
maṛay=o? . . .
put.down=ACT.PST
‘And [. . .] he put down the palanquin **on the pretence** of having to
pee ’ [AK, 2:15]
192. *ro ho=kaṛ* lebu=ki bay=na=ya? *buidh*,
and that=SG.HUM person=PL make=INF=GEN idea
bay=na=ya? *tarika* *socay=na* *maṛe=yo?*
make=INF=GEN method think=INF begin=ACT.PST
‘And He (= God) began to think of **an idea**, of **a way** of making
people.’ [AK, 3:9]
193. . . . *ro gaṛi* ḍe=na=? *samay=jo* *ḍel* *ḍam=ki*.
and (train)car come=INF=GEN time=ADD come arrive=MID.PST
and the **time** for the railway car to come arrived.’ [HJPa:54]
194. *ho=kaṛ=te* sathay=na=? *darkar* *umbodej*.
that=SG.HUM=OBL rest=INF=GEN necessity NEG.QUAL.PRS
‘He (= God) does not need to rest.’ (= does not have a *resting*
necessity.) [HJPa:101]

The distribution of this attributive construction is given in Table 7.9.

Table 7.9: The distribution of the prenominal genitive-case infinitive

(O _{REL} = O _{MAIN})	(Instrument _{REL})	(Genitive _{REL})
	(Locative _{REL})	
(O _{REL} = A _{MAIN})	(Temporal _{REL})	

As Table 7.9 shows, despite the fact that this construction is found in our texts as well as those in Pinnow (1965a), it is not acceptable at all for some speakers and even for those speakers for whom it is acceptable, it has only a very limited distribution.

Prenominals with a direct-case infinitive

This construction functions along the same lines as that described above, except that the infinitive is unmarked for case (= direct case). The following provides a few examples. Its distribution is given in Table 7.10.

195. *musniŋ u=ghay hoy=ki no ɖoli=te ɖoʔɖ=na*
 one.day this=way become=MID.PST Cmpl palanquin=OBL take=INF
bhere u kharīya pahan=te haɖa laʔ=ki.
 time this Kharia priest=OBL pee EMOT=MID.PST
 ‘One day it hapened this way that, **at the time** [they were] to take away the palanquin, the Kharia priest had to pee.’ [AK, 2:8]
196. *ho=kaɖ khoɖi buli=na modhe buŋ*
 that=SG.HUM village.section wander=INF means INST
kharīya=ki=yaʔ jhari habhaw=te erikhudi koŋ
 Kharia=PL=GEN all mannerisms=OBL from.bottom.to.top know
may=sikh=oʔ.
 TOTAL=PERF=ACT.PST
 ‘She had learned through wandering through the village all of the mannerisms of the Kharia inside and out (= **through the in the vil- lage wandering means**’).’ [MT, 1:254]

Table 7.10: The distribution of the prenominal direct-case infinitive

S _{REL} = S _{MAIN}	O _{REL} = O _{MAIN}	A _{REL} = A _{MAIN}	Instrument _{REL}	Genitive _{REL}
S _{REL} = O _{MAIN}	O _{REL} = S _{MAIN}	A _{REL} = S _{MAIN}	Locative _{REL}	
S _{REL} = A _{MAIN}	O _{REL} = A _{MAIN}	A _{REL} = O _{MAIN}	Temporal _{REL}	

Unlike the prenominal construction with a genitive-case infinitive, the prenominal construction with a direct-case infinitive is quite productive and is acceptable to all in the functions given in Table 7.10.

Participles

Two of the following three strategies are the only purely attributive constructions of the language, both of which have been borrowed from Indo-Aryan. Their use is considered by many to be incorrect. The third construction makes use of *-du?* and is still somewhat elusive in nature, although it is highly productive in this function for some speakers. Cf. sections 3.3 and 6.6.2.5 for more discussion on these three forms. In the following, a few simple examples of these three constructions are provided as well as tables showing their distribution.

I. Prenominal construction with “Infinitive + wala”

197. *toblun=te maha maha camke=na=wala siŋkom aw=ki...*
 sky=OBL big REP shine=INF=PTCP star QUAL=MID.PST
 ‘In the sky there were big, *shining stars*. [HJPa:100]

198. *jharkhand=te aw=na=wala lebu=ki ikud jughay milansar*
 Jharkhand=OBL live=INF=PTCP person=PL very much friendly
aw=ta=ki.
 QUAL=MID.PRS=PL
 ‘**The people** *who live in Jharkhand* are very friendly.’

Unfortunately, the data for this construction could only be discussed with two speakers, both from southwestern Simdega District. Table 7.11 shows the distribution for these two speakers.

Table 7.11: The distribution of the prenominal participle =wala

S _{REL} = S _{MAIN}	O _{REL} = O _{MAIN}	A _{REL} = A _{MAIN}	Instrument _{REL}	Genitive _{REL}
S _{REL} = O _{MAIN}	O _{REL} = S _{MAIN}	A _{REL} = S _{MAIN}	Locative _{REL}	Comparative _{MAIN}
S _{REL} = A _{MAIN}	O _{REL} = A _{MAIN}	A _{REL} = O _{MAIN}	Temporal _{REL}	

II. Prenominal construction in *-l* The use of this construction is restricted to contentive morphemes which have been borrowed from Sadri and which end in *-a(y)*. What would be the subject of the corresponding finite clause appears in the genitive, unless it is the modified head. The participle in *-l* directly precedes the head.

199. *muda moṅ brahman* ho=kaṛ=a? daru=te taṅa-l
 but one brahman that=SG.HUM=GEN tree=OBL hang-PTCP

janew=te yo=yo?
 holy.thread=OBL see=ACT.PST

‘But a Brahman_i saw **the holy thread** which he_j had hung on the tree.’ [AK, 2:4]

Further examples:

153. *melay* ‘leave’ mela-l **thayla** ‘a forgotten (= left) bag’
bihay ‘marry’ biha-l **lebu** ‘a married man’
batay ‘tell’ bata-l **kahani** ‘a story which has been told’

Note that, with the exception of $A_{REL} = A_{MAIN}$, which one speaker rejected (out of a total of four questioned with respect to this construction), the participle in *-l* is entirely free in its distribution, as Table 7.12 shows.

Table 7.12: The distribution of the prenominal participle *-l*

$S_{REL} = S_{MAIN}$	$O_{REL} = O_{MAIN}$	$(A_{REL} = A_{MAIN})$	Instrument _{REL}	Genitive _{REL}
$S_{REL} = O_{MAIN}$	$O_{REL} = S_{MAIN}$	$A_{REL} = S_{MAIN}$	Locative _{REL}	Comparative _{MAIN}
$S_{REL} = A_{MAIN}$	$O_{REL} = A_{MAIN}$	$A_{REL} = O_{MAIN}$	Temporal _{REL}	

III. Prenominal construction in *-ḍu?* All examples for this construction were elicited, as it occurs nowhere in our corpus nor in the that part of Pinnow (1965a) which we have so far analyzed in this respect. One speaker said that it is especially common of the older generation. The following presents a few examples.

200. *dilli* col-ḍu? **lebu** *ute* *aw=ta*
 Delhi go-PTCP? man here live=MID.PRS
 ‘**The man** who went / is going / will go to Delhi lives here.’

201. ho=ki=ya? *likha-ḍu?* **phonten=te** *kuy=o²j*.
 that=PL=GEN write-PTCP? pen=OBL find=ACT.PST.1SG
 ‘I found **the pen** that they had written / write with.’

Its distribution however, does not give any hints as to why it is so seldomly heard or found in texts, as it is highly productive, at least in principle, as Table 7.13 shows.

Table 7.13: The distribution of the prenominal “participle” *-qu?*

$S_{REL} = S_{MAIN}$	$O_{REL} = O_{MAIN}$	$A_{REL} = A_{MAIN}$	Instrument _{REL}	Genitive _{REL}
$S_{REL} = O_{MAIN}$	$O_{REL} = S_{MAIN}$	$A_{REL} = S_{MAIN}$	Locative _{REL}	Comparative _{MAIN}
$S_{REL} = A_{MAIN}$	$O_{REL} = A_{MAIN}$	$A_{REL} = O_{MAIN}$	Temporal _{REL}	

7.6.3 *Circumnominal clauses*

In what at first glance would appear to be circumnominal attributive clauses, the attributive clause has the same form as a main clause and the head appears in the attributive clause marked according to its function there. The head is not repeated in the main clause.

202. *biru* *ḍeb=na* *ḍeb=na* *ba? runku?b* *ro?=ki*
 mountain ascend=INF REP “rice” spill=MID.PST
kay=e=bar!
 lift=ACT.IRR=2HON
 ‘Pick up (HON) **the rice which spilled while you were walking up the hill!**’ [MT, 1:40]
203. *babu*, *kimin=ki* *jahā* *jo?=na* *cij* *dho?=si?=pe*
 child daughter.in.law=PL INDEF.NHUM eat=INF thing grab=PERF=2PL
gur=na, . . .
 fall=MID.IRR
 ‘Child, **some of the food (= things to eat) which you daughters-in-law have taken** will fall, [MS, 1:44]

The distribution of this construction is given in Table 7.14.

Table 7.14: The distribution of circumnominal clauses

	$(O_{REL} = O_{MAIN})$	$A_{REL} = A_{MAIN}$	Instrument _{REL}
$S_{REL} = O_{MAIN}$	$O_{REL} = S_{MAIN}$	$A_{REL} = S_{MAIN}$	Locative _{REL}
$S_{REL} = A_{MAIN}$	$O_{REL} = A_{MAIN}$	$A_{REL} = O_{MAIN}$	Temporal _{REL}

What is remarkable about this distribution is that, despite its ability to be used with virtually any head other than the genitive, the comparative, or $S_{REL} = S_{MAIN}$ (the latter undoubtedly due to the fact that an interpretation as two coordinated clauses would be preferred), this construction is almost never used in actual speech and our data contain only these two non-elicited examples, both of which incidentally refer to the same incident in traditional Kharia folklore, although from two different speakers. Thus,

although this construction is potentially very productive, it seems to have almost completely fallen into disuse. As one speaker noted, laughing, when he was presented with a similar example, “It’s OK, but no one talks like that.” While the other speakers were not as skeptical, the use of this construction is certainly very rare. As late as the early 1980’s, however, Malhotra was still able to note its use (cf. e.g. Malhotra, 1982: 315).

The status of these constructions as circumnominal propositional attributes is somewhat dubious, as they could also be interpreted as two finite clauses, where one argument (= the “head”) is coreferentially deleted in the second clause. We will now take a closer look at this topic.

Examples such as (202) above are generally also interpretable as two main clauses, depending on intonation. This is possible because Kharia does not make use of any type of syntactic pivots (i.e., neither S/A nor S/O) in coreferential deletion (7.1.2). Thus (202) could also be interpreted as a simple juxtaposition: “[While] you were walking up the hill, rice spilled. Pick [it] up.” The following presents a further example.

204. *in lebu=ki=te yo=yoʔj peʔ nokh=oʔ=ki.*
 1SG person=PL=OBL see=ACT.PST.1SG rice eat=ACT.PST=PL
 ‘The people I saw ate rice.’

This example can also be interpreted as “I saw the people and they ate rice.” However, there are slight differences which still must be investigated in greater detail. For example, in (204) the interpretation of the first clause as an attributive clause is preferred. If the interpretation is that of two independent clauses, the conjunction *ro* ‘and’ is invariably inserted.

Furthermore, these “circumnominals” cannot carry case marking, which would argue against treating them as complex Case-syntagmas. In other words, although *in lebukite yoyoʔj* in (204) may appear to be the subject of the clause, we cannot mark the entire phrase, e.g., with the oblique marker *=te* in object function.

205. **lebu=ki=te yo=yoʔj=te gil=oʔ.*
 person=PL=OBL see=ACT.PST.1SG=OBL hit=ACT.PST
 ‘S/he hit the people I saw.’

This is due to the general prohibition on combining case with markers of the TAM/PERSON-syntagma, discussed in Chapter 4, as a clausal unit can only be EITHER a Case-syntagma OR a TAM/PERSON-syntagma, but not both simultaneously. This also strongly argues against considering this a

“circumnominal” attribute and suggests that these are simply two juxtaposed clauses.

There is however at least one argument against treating “circumnominals” as juxtaposed clauses: Not all instances of this construction are interpretable as two main clauses. As we saw above, (204) can be interpreted as two main clauses, but the conjunction *ro* ‘and’ is then invariably used. Another example:

206. *lebu* *ijɪ=te* *yo=sikh=oʔ* *goʔj* *goʔd=ki*.
 person 1SG=OBL see=PERF=ACT.PST die C:TEL=MID.PST
 a. ‘The person who saw me died.’
 b. ‘The person saw me and died.’

For whatever reason,²⁵ only one speaker accepted the meaning given in (206)b, while one rejected it with this meaning. Both speakers, however, as well as a third speaker, all readily accepted it with the meaning given in (206)a, i.e., the circumnominal construction.

What is important for our discussion here, however, is that there clearly is a difference for speakers between these two interpretations, and there are at least some instances of circumnominal clauses which resist an interpretation as two coordinated clauses and *vice versa*.

There is also the question of intonation: While there seem to be different intonational patterns, this is at the moment purely impressionistic and we will refrain from making any further comments here. The main difficulty involved is that the data contain so few non-elicited examples, all others are to a certain extent “artificial” Further research is necessary.

7.6.4 *Postnominals*

Variations of the two correlative constructions are used postnominally in a slightly modified form, often with non-restrictive clauses. As we do not have detailed information on these constructions, the following merely presents a few examples from the corpus. These constructions await further study.

207. *ro* ***brahman=ki***, *je* *tama* *pujapaṭh* *karay=te=may*,
 and Brahman=PL CREL now sacrifice do=ACT.PRS=3PL

²⁵ Perhaps as this example is semantically somewhat far-fetched?

ho=ki ḍoli=te goʔ=na laʔ=ki=may.
 that=PL palanquin=OBL carry=INF IPFV=MID.PST=3PL

‘And the **Brahmans**, who now do sacrifices, they used to carry the palanquin.’ [AK, 2:7]

208. haḍa=na thoṅ ho=kaṛ gam=oʔ lebu=ki=te no
 pee=INF PURP that=SG.HUM say=ACT.PST person=PL=OBL Cmpl

“tama katiʔj deri goʔ=na melay[=e]=pe ro u=ghay
 now some time carry=INF leave=ACT.IRR=2PL and this=way

thāṛo=te melay[=e]=pe no jahā boʔ=te
 place=OBL leave=ACT.IRR=2PL Cmpl CREL place=OBL

daru=ki ayiʔj.”

tree=PL QUAL.PRS

‘In order to pee, he said to the people “Stop carrying [the palanquin] for a little while, now, and stop in a kind of place where there are trees.”’ [AK, 2:9]

209. muda moṅ dhāgar kongher, ber merom gupa=na laʔ=ki,
 but one servant boy who goat shepherd=INF IPFV=MID.PST

ho=kaṛ iḍib=te aḍi=yaʔ bāṭa kuda kolon=te
 that=SG.HUM night=OBL ANAPH=GEN share millet bread=OBL

sob=jeʔ um joʔ=na laʔ=ki ro haʔdo=jeʔ
 all=SG.NHUM NEG eat=INF IPFV=MID.PST and half=SG.NHUM

un=na laʔ=ki.

place=INF IPFV=MID.PST

‘But **one servant boy**, who used to tend the goats, he didn’t use to finish all his share of the millet bread at night and he placed [down] half of it.’ [BB, 1:13]

7.6.5 The scale of relativization

The data available on the distribution of the various constructions is brought together in Table 7.15, having abstracted away somewhat from the individual distributions presented in the previous pages by grouping S and A together under the heading “Subject” and O as “Object” The data here refer to the attributive clause only, unless otherwise noted.

Table 7.15: The scale of relativization in Kharia

<i>Subject</i>	<i>Object</i>	<i>Temporal_{REL}</i>	<i>Locative_{REL}</i>	<i>Instrument_{REL}</i>	<i>Genitive_{REL}</i>	<i>Comparative_{MAN}</i>
<i>je</i>	<i>je</i>	<i>je</i>	<i>je</i>	<i>je</i>	(<i>je</i>)	<i>je</i>
Q	(Q)	Q	Q	Q	Q	Q
<i>wala</i>	<i>wala</i>	<i>wala</i>	<i>wala</i>	<i>wala</i>	<i>wala</i>	<i>wala</i>
<i>-du?</i>	<i>-du?</i>	<i>-du?</i>	<i>-du?</i>	<i>-du?</i>	<i>-du?</i>	<i>-du?</i>
<i>-l</i>	<i>-l</i>	<i>-l</i>	<i>-l</i>	<i>-l</i>	<i>-l</i>	<i>-l</i>
((*)A _{REL} = A _{MAIN})						
DIR-INF	DIR-INF	DIR-INF	DIR-INF	DIR-INF	DIR-INF	—
Masdar	Masdar	Masdar	Masdar	Masdar	—	Masdar
Circum- nominal	Circum- nominal	Circum- nominal	Circum- nominal	Circum- nominal	—	—
(*S _{REL} = S _{MAIN})	(*O _{REL} = O _{MAIN})					
(FF/PF)	(FF/PF)	FF/PF	(FF/PF)	—	—	FF/PF
—	(GEN-INF)	(GEN-INF)	(GEN-INF)	(GEN-INF)	(GEN-INF)	—

DIR-INF—prenominal with direct-case infinitive

FF/PF—prenominal with a fully finite or partially finite predicate

GEN-INF—prenominal with genitive-case infinitive

As can be seen in the table, the correlative and participial constructions are the potentially most productive strategies (although not necessarily the most common) and can be used for propositional attributes whose head is in any of the functions investigated here. Prenominals with a direct-case infinitive or the masdar are equally productive and can be used for any head except a comparative (direct-case infinitive) or a genitive (masdar). Circumnominals are slightly less productive (although much less common), as are prenominals with a fully or partially finite predicate, although these latter two are far more common than the circumnominals in texts. Finally, the prenominal construction with a genitive case infinitive is only marginally acceptable.

Leaving aside the postnominal structures, which as we have seen largely reduplicate the correlatives, which are certainly due to Indo-Aryan influence, we can classify propositional attributes in Kharia as in Diagram 7.1.

Circumnominals	Prenominals	
Fully finite	Partially finite	Non-finite
Propositional attribute (generally) outside the semantic base of a Case-syntagma	?	Propositional attribute within the semantic base of a Case-syntagma
Potentially fully independent	Subordinated—degree of dependency increases from left to right	

Diagram 7.1: Schematic overview of various propositional attributes

Constructions with fully finite forms overlap both the circumnominal and prenominal constructions, whereas all partially finite and non-finite constructions are prenominal. Furthermore, those constructions which contain a fully finite TAM/PERSON-syntagma are almost always external to the semantic base of a Case-syntagma, i.e., they are simply juxtaposed to the Case-syntagma. Non-finite constructions, on the other hand, are always internal to the semantic base of a Case-syntagma. The status of partially finite forms is unclear in this respect, due to lack of data.

Finally, the degree of potential independence directly correlates with the degree of finiteness of the construction: Whereas circumnominal and fully finite prenominal clauses are potentially fully independent, the non-finite constructions are dependent structures and are at best only marginally found as independent clauses, such as the use of the masdar with *umbo?* in negation (6.7.3). Partially finite prenominals occupy an intermediate position: As they lack person marking, they cannot stand as the predicate of a main clause. They can, however, appear in complex predicates of the type described in 6.6.1.

The prenominal constructions also demonstrate the full continuum of morphological finiteness, from fully finite forms to partially finite and non-finite forms. This is illustrated schematically in Diagram 7.2.

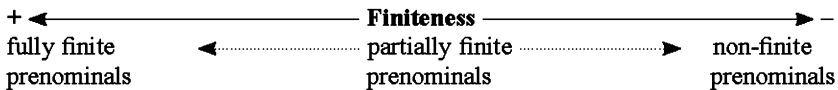


Diagram 7.2: Continuum of finiteness in prenominal propositional attribution

Finally, the data presented in this section allow us to speculate as to the development of prenominal propositional attributes in Kharia. Note

that there are other, non-Munda, Austro-Asiatic languages such as T'in (Filbek, 1976) which possess what appear to be circumnominal propositional attributes, like Kharia. Although this is highly speculative at the moment and requires further research, it would not be unreasonable to assume that the development of prenominal constructions followed the development outlined schematically in Diagram 7.3.

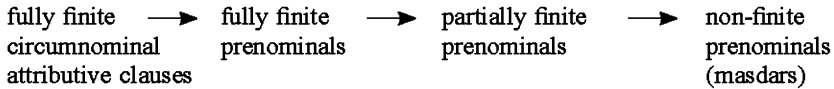


Diagram 7.3: Suggested historical development of prenominals from circumnominals

In this suggested development, fully independent “circumnominals” were simply independent clauses, one element of which was coreferential with an element of the following clause, which was hence omitted there. At this stage, we are probably still dealing with a predicate-medial language.

As our hypothetical proto-language gradually shifted to predicate-final (or, more generally, head-final, cf. Donegan & Stampe, 2004), the “circumnominal” construction gradually evolved into the prenominal construction with a fully finite predicate. This development merely consisted of what may be considered the “gapping” of the lexical head in the attributive clause: Instead of the lexical head appearing in the attributive clause, as in the “circumnominal” construction, it now appears in the main clause, marked for its function there. The “circumnominal” construction, however, continued—and continues—to be used, although gradually losing ground to the newer construction.

In the course of time, person-marking became optional in the preposed construction, especially in the third-persons, as the attributive “clause” became more integrated into the semantic base of a Case-syntagma, resulting in the partially finite prenominal construction. Finally, not only person marking but all TAM marking which is obligatory for a finite predicate became optional, resulting in the non-finite or masdar construction. This later development was accompanied by the full integration of the propositional attribute into the semantic base of a Case-syntagma.²⁶

This suggested development is of course purely speculative at the moment, although it would seem to account for the data quite well, as

²⁶ As noted above, the correlative constructions can safely be assumed to derive from contact with Indo-Aryan. As such, they are a later development.

well as the fact that the non-finite prenominals (i.e., those with the *masdar*) are now so common and productive. Perhaps this suggested development can be refined further as research into propositional attribution in other Munda languages progresses.

7.7 INFORMATION STRUCTURE—TOPIC, FOCUS AND CONSTITUENT ORDER

The order of constituents in a Kharia sentence is independent of the function of the respective constituent, i.e., whether it refers to the subject, object or an adjunct. The order of these elements is determined purely by their pragmatic status, with topical and/or activated elements preceding focal/inactive elements. Thus, while there is a tendency for Case-syntagmas referring to subjects to appear first in the clause and those referring to the object to be the last constituent preceding the TAM/PERSON-syntagma, this is not necessarily the case and any ordering of the constituents of the clause is possible.

The sentence often begins with a “scene-setting topic” which provides the background or scene-setting information for the remaining sentence. Alternatively, this slot may contain focal information, as in argument focus. It is often followed by a brief pause. This slot is then followed by the remaining sentential constituents and the TAM/PERSON-syntagma. The order of these remaining constituents is as follows: Topical elements appear first, followed by focal elements, usually with the TAM/PERSON-syntagma as the last element. Occasionally we also find a topical or focal constituent following the TAM/PERSON-syntagma. Thus the schematic structure of a Kharia sentence is as in Table 7.16.

Table 7.16: Pragmatic structure of the Kharia sentence

	<i>Prefield</i> (optional)		TAM/PERSON- syntagma	<i>Postfield</i> (optional)
Scene-setting Topic or Focused Case- syntagma	Topical Case- syntagmas	Focal Case- syntagmas		Topical or focal Case-syntagma

This structure will be elaborated in the following sections, divided into predicate focus or topic-comment structure (7.7.1), argument focus (7.7.2), and sentence focus (7.7.3). As we shall see, however, this division is unnecessary for Kharia, at least from a morphosyntactic perspective, as all three focus-types have the same basic structure. 7.7.4 then presents

a brief description of the three most common “pragmatic” enclitics and their functions.

7.7.1 *Predicate focus or Topic-Comment*

The topic-comment structure, referred to by Lambrecht (1994) as the predicate-focus structure, may be considered the “unmarked” structure in which a statement is made about the topic of the sentence. In Kharia this structure is generally one in which a topic, perhaps preceded by a “scene-setting topic”, appears at the beginning of the clause and is followed by the focus, which provides information about this topic.

The topic is typically a Case-syntagma which refers to the subject of the clause, although this is not necessarily the case, and the focus generally consists of the TAM/PERSON-syntagma and perhaps a Case-syntagma referring to the object and one or more adjuncts. The topic is typically “activated”, i.e., its identity is typically already established, and is usually predictable, whereas that of the focus is not, although the focus may also contain topical elements (cf. Lambrecht, 1994).

Consider the following example from the beginning of a story in which the topical and/or activated elements of each sentence are given in **bold-face** and the focus is *italic* in the English translation. As (210)a is the opening line, this is an example of what Lambrecht (1994) refers to as a presentational construction and is hence an example of sentence focus, with no topic. This will be dealt with in 7.7.3 below.

210. a. **moŋ** gupa lebu aw=ki.
 one guard person QUAL=MID.PST
 ‘*There [once] was a shepherd.*’ [RD, 1:1–3]
- b. **ho=kaŋ** diyo=ga merom gupa=na co=na la?=ki.
 that=SG.HUM daily=FOC goat guard=INF go=INF IPFV=MID.PST
 ‘*He used to watch over goats daily.*’
- c. **ho=kaŋ=a?** iku’d jughay **bheŋi** merom aw=ki=may.
 that=SG.HUM=GEN very much sheep goat QUAL=MID.PST=3PL
 ‘*He had very many sheep and goats.*’
- d. **ho=kaŋ** d̥isa? d̥isa? kho?tay=jō merom gupa = na
 that=SG.HUM far REP up.to=ADD goat guard=INF
la? = kho?
 IPFV=PST.II
 ‘*He used to tend the goats, even going (= up to) very far.*’

Notice that in (210)b, c and d the topic, *mon gupa lebu*, is represented by a proform as it has been activated in (210)a. Also, as it is the topic, it appears at the beginning of the sentence, preceding the focus.

Much more commonly, activated and/or topical elements are simply left unexpressed, as in the following example from the same story ([RD, 1:15–16]). The symbol \emptyset_1 refers to the shepherd, \emptyset_2 to the livestock. Note that any topical element may be omitted, regardless of whether it refers to the subject, object or another entity (e.g., recipient in triadic predicates such as ‘give’). Note also that in (211)b the object of the clause is the livestock, i.e., *lam-lam* is transitive, despite the English translation.

211. a. α^2b *le²bqom₁* leme²d₁ tay jayqim=ki tay \emptyset_1
 now boss sleep ABL wake.up=MID.PST then
merom=qom=ki=te₂ han=ti²j u=ti²j yo=te, *la² \emptyset_2*
 goat=3POSS=PL=OBL that=side this=side see=ACT.PRS then
u=ti²j=jo um yo=ta.
 this=side=ADD NEG see=MID.PST
 ‘Now **the boss** woke up from his sleep and [\emptyset_1] looks for his goats in this and that direction but even in this direction [\emptyset_2] are not seen.’
- b. *bhere \emptyset_1 \emptyset_2* han go²a=ga lam-lam khor=te.
 time that whole=FOC search-RDP ITER=ACT.PRS
 ‘[At that] time [\emptyset_1] looks all around that whole [area] [for \emptyset_2].’

In addition, α^2b in (211)a and [*ho*]bhere ‘[that] time’ in (211)b also show that the temporal scene-setting topic appears first in the clause, as it provides the background information for what is to come. As the following example shows (from the same story, [RD, 1:19–21]), especially (212)a, other types of adjuncts can also occupy this position and serve as scene-setters, not just temporal and locative adjuncts. Note that the topic, *ho=ka²=te* ‘him’, is only overt in (212)a and is not mentioned again in the following lines.

212. a. *du²kho bu² ho=ka²=te* jiyom=te jiyom um dho?
 SORROW INST that=SG.HUM=OBL life=OBL life NEG grasp

qom=ki.

PASS=MID.PST

'Due to his sorrow, he *could not bear to live any more.*'²⁷

- b. *bhere* murjhay=kon gur go³q=ki.
 time hang.the.head.low=SEQ fall C:TEL=MID.PST
 '[At that] time, [he] hung his head low and dropped to the ground (= fell).'

- c. *ro* ho=te=ga go³j go³q=ki.
 and that=OBL(= 'there')=FOC die C:TEL=MID.PST
 'And, right there, [he] died.'

The following example shows that constituents other than merely those referring to the subject or the direct object may also be omitted, in this case, those referring both to the direct object of the clause and to the recipient. The inclusion of a Case-syntagma referring to the subject, *ijn* '1SG', is probably due here to the fact that the topic has now been changed, a kind of "contrastive topic" in Lambrecht's terminology.

213. *i lam=te=m? la? mane ijn ter[=e]=ijn lekin*
 what want=ACT.PRS=2SG then umh 1SG give[=ACT.IRR]=1SG but

ijn=te jan abu tar=e=m.
 1SG=OBL life NEG.MOD kill=ACT.IRR=2SG
 'What do you want? Then, umh, I will give [it to you], but don't kill me.' [MS, 1:64]

In the following example, a Case-syntagma referring to the subject follows that referring to the object, as this is more active or more topical. Here, the ant is telling the grasshopper that she cannot give him any sugar, since her sisters have worked hard gathering it. The sisters are mentioned here for the first time in the story and are hence less topical than the sugar, which the grasshopper has just asked for.

214. *u cini ijn=a? bohina=na=jn=ki thuray=si?=may.*
 this sugar 1SG=GEN sister=1POSS=1SG=PL gather=PERF=3PL
 'This sugar my sisters have gathered.' [TK, 2:42]

²⁷ This is an idiomatic expression and cannot be analyzed further. Literally it translates as 'Through sorrow, to him, in life, life was not grasped.'

Also, although the TAM/PERSON-syntagma is usually in clause-final position, this is not obligatory. Consider the following example, in which the activated and topical sons are first omitted (represented by “ \emptyset_2 ”, as their identity is clear from the preceding text (not given here)). They are then referred to in the second instance, after the TAM/PERSON-syntagma, but only with respect to their number (= ‘all of them’).

215. *aba=dom=kiyar*₁ \emptyset_2 um=kiyar batay=o? no i jinis
 father=3POSS=HON NEG=HON tell=ACT.PST Cmpl what animal
 heke, i jhãut heke *lekin, muda*, \emptyset_1 \emptyset_2
 QUAL.PRS what animal QUAL.PRS but but
 d̪aŋ goʰh=o?, souʼb=te=ga₂
 send C:TEL=ACT.PST all=OBL=FOC
 ‘**Their father** didn’t tell [*them*] which animal it is, which animal it is, but [*he*] sent [*them*] off, all [*of them*].’ [AK, 1:10]

Especially common is the placement of the goal with contentive morphemes of motion after the TAM/PERSON-syntagma, as in the following example:

216. *raylogarh tay muʔ=kon purkha=ki* dhirom dhirom
 Railogarh ABL emerge=SEQ ancestor=PL slowly REP
 d̪el=ki=may d̪eld̪aʔpur.
 come=MID.PST=3PL Delhi
 ‘Having left Railogarh, **the ancestors** gradually came to Delhi.’ [MT, 1:106]

The status of the post-TAM/PERSON-syntagma constituent in (216) differs from that in (215) in that the post-TAM/PERSON-syntagma element in (216) is non-activated focal information. Thus, the constituent following the TAM/PERSON-syntagma in the postfield may be either topical or focal.

Although goals with contentive morphemes of motion often appear after the TAM/PERSON-syntagma, this is not obligatory, as the following example shows. The difference between the two in terms of their respective pragmatic statuses is as yet unclear, as both structures contain new, focused information.

217. *khariya=ki* darjilin=jo d̪am=sikh=oʔ=ki.
 Kharia=PL Darjeeling=ADD arrive=PERF=ACT.PST=PL
 ‘**The Kharia** also came to Darjeeling.’ [MT, 1:52]

As noted above, the first Case-syntagma of the clause is commonly a scene-setting unit and can also introduce a new topic. This unit may often be referred to again in the clause, often by a proform. Both the new topic as well as the anaphoric reference to this element (and other topical information) are given in **bold face** and the focus is *italic* in the following.

218. *ho=ki soub* pujapaṭh=aʔ ghaḍ d̪oko goḍ=ki=may,
 that=PL all sacrifice=GEN for sit.down C.TEL=MID.PST=3PL
 cerocagordi, *pahan, khariya pahan, ho=kaṛ* d̪el=ki,
 on.all.four.sides priest Kharia priest that=SG.HUM come=MID.PST
pujapaṭh karay=na.
 sacrifice do=INF
 ‘**They all sat down for the sacrifice, on all four sides, the priest, the Kharia priest, he came, to do the sacrifice.**’ [AK, 2:18]

219. *lebu=ki adha=ki* gor aʔj=may, *adha=ki* telsāwār
 person=PL half=PL fair QUAL.PRS=3PL half=PL dark
adha=ki moghere aʔj=may.
 half=PL black QUAL.PRS=3PL
 ‘**The people, half are fair, half are dark, half [sic!] are black.**’

u khariya=ʔ hoʔ=ki, seṅ=ko loʔkha=yaʔ *hoʔ* bay=na
 this Kharia=GEN house=PL first=CNTR dirt=GEN house build=INF
laʔ=ki=may, lekin, muda, tama simēt, iṭā, soreṅ
 IPFV=MID.PST=3PL but but now cement brick stone
mesa=ke hoʔ bay=te=may.
 mix=SEQ house build=ACT.PRS=3PL
 ‘**These Kharia’s houses, they earlier (= first) used to make houses of dirt, but now they mix cement, bricks and stone and make houses.**’
 [AK, 5:7f.]

7.7.2 *Argument focus*

Sentences in which a single constituent is within the scope of the focus function along the same lines as topic-comment structures in that topical elements typically appear first, followed by the focused constituent, which typically precedes the TAM/PERSON-syntagma. Consider the following example from [AK, 5:24–5].

220. a. [Since the time Jharkhand became independent, therefore they speak Kharia in school as well.]
 b. *lekin sej* umay kayam=na laʔ=ki, *sej=ko* hindi
 but first NEG.3PL speak=INF IPFV=MID.PST first=CNTR Hindi
kayam=na laʔ=ki=may, oḍo [ʔ] nagpuri.
 speak=INF IPFV=MID.PST=3PL and Nagpuri
 ‘But **at first [they] didn’t speak [Kharia], at first [they] spoke Hindi, and Nagpuri.**’

In (220)b *hindi*, as well as the postposed *nagpuri*, is focused—the fact that the children speak *something* at school can be considered topical information as it has just been stated that they do not speak Kharia (which is topical from the preceding lines and is hence omitted in (220)b). Thus in (220)b it is *hindi* which is focused, as well as the postposed *nagpuri*, which however is entirely outside of the clause. Here as well we find the usual order Topical > Focal.

7.7.3 *Sentence focus*

In sentence-focus constructions, including presentational structures, the order is typically as follows:

221. Scene-setting adjunct—Case-syntagma referring to the subject—
 TAM/PERSON-syntagma

All presentational structures that we have examined are intransitive. This was seen above in (210)a and is also demonstrated by the following example.

222. akhāṛ kinir=te moṅ soloʔ aw=na laʔ=ki.
 deep forest=OBL one dog live=INF IPFV=MID.PST
 ‘In a deep forest a dog once lived.’ [BB, 3:1]

Another quite common structure consists of a scene-setting topic such as *meson* ‘once’ with a predicate containing *hoy=ki no* [become.MID.PST Cmpl] ‘it happened that’ plus the structure in (221). This is especially common in narratives when there is a change of scene or an entirely new event is to take place.

223. *meson* u=ghay hoy=ki no d̪heir dino tay barkha
 once this=way become=MID.PST Cmpl many day ABL rain
 um d̪el=ki.
 NEG come=MID.PST
 ‘It **once** happened thus that for many days rain did not fall
 (= come).’ [TK, 1:5]

In sum, there appear to be no major structural differences between the various focus-types discussed in Lambrecht (1994)—predicate-focus, argument-focus and sentence-focus, although especially these two latter types require further study. In the first two types we find topical elements preceding focal elements, although the first slot, at the beginning of the sentence, may be either topical or focal. Sentence-focus makes use of the same strategy—scene-setting information appears at the beginning of the clause, followed by a Case-syntagma referring to the subject and the TAM/PERSON-syntagma.

7.7.4 *The pragmatic markers =ga, =ko and =jo*

In this section the three most common pragmatic enclitics are discussed and an approximation of their functions is offered. Other pragmatic markers, such as the “focal” markers =aʔ or roʔ, will not be discussed due to the paucity of materials. These await further research. For further information on these, see the respective entries in Peterson (2009).

All three markers have in common that they deal with INTERSENTENTIAL pragmatic structures. At the sentence level we have predicate focus, argument focus and sentence focus, discussed in the previous sections, while the pragmatic markers discussed here have to do with expectations from one sentence to another, e.g., topic (dis)continuity, contrastive topics, focus, etc. However the intrasentential pragmatic structure of the clause is not affected by the presence of these markers.

The additive-focus marker =jo 'ADD'

=jo is the most straight-forward of the three particles. It denotes inclusive or additive focus, hence the gloss 'ADD' for "additive", corresponding quite nicely to the English words 'also' or 'even'. It differs from the other two pragmatic markers in this respect, as these latter markers are exclusive or restrictive focal markers. The following presents a simple, typical example.

After saying that the Kharia had travelled to Burma (via Nepal), Sikkhim, Arunachal Pradesh, Nagaland, Manipur, Meghalaya, Tripura and Mizoram (all in northeastern India), the speaker adds:

224. *khariya=ki darjiliŋ=jo dam=sikh=o?=ki.*
 Kharia=PL Darjeeling=ADD arrive=PERF=ACT.PST=PL
 'The Kharia *also* came to *Darjeeling*.' [MT, 1:52]

=jo is also found in combination with interrogative and other proforms in negated clauses to denote 'nobody', 'nothing', 'never', e.g., *behar* 'who', *behar=jo umbo?* [who=ADD NEG] 'no one' (cf. 5.7 for further examples).

The restrictive focus markers =ko 'CNTR' and =ga 'FOC'

=ko and =ga differ from =jo in that they are not additive or inclusive focus markers but rather restrictive or exclusive. There is no real English translation for the two, the closest translational equivalent being perhaps 'just', 'exactly' or 'only'. Much research on these two enclitics remains to be done and the following is at best a very general guide to their use.

In general, the function of these two enclitics is very similar, and neither is ever obligatory, except =ga in depictive secondary predicates (cf. 7.5.1.6).²⁸ Otherwise, their use is entirely facultative and their frequency can vary greatly from one speaker to another and from one text genre to another. The most common use of these two enclitics is to denote a contrastive topic or focus from one clause to the next. The difference between the two is largely that whereas =ko generally has a strong "contrastive" flavor to it, serving to highlight one entity instead of a single other entity (i.e., one of two), =ga denotes that the new topic / focus is one of any number of possibilities.

²⁸ For ease of presentation, the use of =ga in depictive secondary predications is excluded from the following discussion, as it may be more appropriate to consider it a depictive marker in that construction.

In the following example, Porha, the third of nine sons, has just spoken (text given here in English).

225. ‘Porha said “Maybe, umh, he said to kill a deer, a deer.”’ [AK, 1:13]

226. *la?* *sou²b=ga* *bhai=ki* *ho* *khajar* *tar=na* *ghaḍ* *juda*
 then all=FOC brother=PL that deer kill=INF PURP separate

juda *mu?²=ki=may,* *kinir=te.*
 REP emerge=MID.PST=3PL forest=OBL

‘Then *all* the brothers set out separately to kill that deer, into the forest.’ [AK, 1:15]

The use of =*ga* here on *sou²b* emphasizes that now all nine brothers, not just Porha or potentially any one of the other eight, are meant.

Consider now the following three consecutive sentences from a story [AK, 1:68–70], each of which contains an element marked for =*ga*. Here, the speaker (the father) is telling his nine sons that the animal they found in their bundles after hunting is the totem of their future respective families:

227. *je* *je* *yo=sikh=o?²=pe* *ho=je[ʔ]=ga* *am=p=a?*
 CREL REP see=PERF=ACT.PST=2PL that=SG.NHUM=FOC 2=2PL=GEN

kuṭum *heke.*
 family QUAL.PRS

‘Whatever you saw, *that* [*and nothing else*] is your [future] family.’

228. *ro* *tama* *am=pe* *u* *nāw* *kuṭum=te=ga* *sadi* *biha*
 and now 2=2PL this nine family=OBL=FOC marry marry

kerson=na=pe *ro* *bes-bo?*, *baru-bo?* *aw=na=pe.*
 marry=MID.IRR=2PL and good-INTENS good-INTENS QUAL=MID.IRR=2PL

‘And now you will marry *in only these nine families* [*and no others*] and you will be very happy.’

The story then ends with a topic change, marked by =*ga*:

229. *u=je[ʔ]=ga* *heke* *khariya* *gotar=a?* *choṭka* *kahani.*
 this=SG.NHUM=FOC QUAL.PRS Kharia clan=GEN small story

‘And *this* is the short history of the Kharia clan.’

In contrast, consider now the use of =*ko* in the following examples.

233. *hoʔ=te* *aba=ɲ* *ayiʔj*, *ayu=ɲ=ko* *col=ki*
 house=OBL father=1SG QUAL.PRS mother=1SG=CNTR go=MD.PST
memon=ga *siray=ki*, ...
 year=FOC expire=MD.PST
 ‘My father is at home, *but my mother* died last year (= the year [that] went), [AK, 4:4]

The use of =*ko* on *ayuɲ* denotes the contrast between the father, who is still alive and living at home, and the mother, who had passed away in the previous year.

As the following example shows, =*ko* can also anticipatorily mark the first member of a contrastive pair.

234. *suru=te=ko* *kayam=na* *umay* *lam=te* *lekin ab*
 beginning=OBL=CNTR speak=INF NEG.3PL seek=ACT.PRS but now
ina no *iskul* *boʔ=ki=te=jo* *ab* *khar̥iya=yaʔ* *paʔhai*
 because school place=PL=OBL=ADD now Kharia=GEN teaching
suru *hoy=taʔj*.
 beginning become=MD.PROG
 ‘*In the beginning*, they didn’t (= “don’t”) want to speak [Kharia] but now because, at schools, too, the teaching of Kharia is beginning.’ [AK, 5:23]

Often, however, the notion of “contrast” is less tangible, as in the following example.

235. ‘Then he asked his younger brother “Beghma, did you drink water?”’ [AK, 1:31]

236. *la[ʔ]* *beghma* *gam=o[ʔ]* “*ɪɲ=ko* *ɖaʔ=ga* *um=ɪɲ*
 then Beghma say=ACT.PST 1SG=CNTR water=FOC NEG=1SG
kui=siʔ. *laʔ* *i=ghay* *ud[=e]=ɪɲ* *tayʔʔ*”
 find=PERF then what=way drink=ACT.IRR=1SG then
 ‘Then Beghma said “I have not found water. Then how should I drink?”’ [AK, 1:32]

There is no real contrast here. Nevertheless, the speaker of the second sentence, Beghma, is clearly “emphasizing” that he, unlike perhaps his brother who asked him this question, did not find water. As such, he appears to be setting himself off as one out of a group of two.

Occasionally, however, there is no real sense of contrastiveness or even “emphasis” noticeable from the text nor even a change of context. For example, in the following example the speaker has just mentioned his native village *salḏaʔ* ‘Saldega’ and then repeats it, marked for =*ko*, in the following sentence (from [AK, 5:1–2]).

237. *ɨn=aʔ kʰoʀi=yaʔ ɲimi salḏaʔ heke.*
 1SG=GEN village=GEN name Saldega QUAL.PRS
 ‘The name of my village is Saldega.’

238. *salḏaʔ=ko ikuḏ maha kʰoʀi heke.*
 Saldega=CNTR very big village QUAL.PRS
 ‘Saldega is a very big village.’

It is, however, possible that the speaker here would like to stress that, contrary to what the listener may be thinking, Saldega is a large village, i.e., ‘But Saldega is a very big village.’ More research is necessary here.

Finally, =*ko* is obligatory in a small number of fixed expressions, such as *laʔ=ko* [then=CNTR] ‘but’, *cahe=ko* [or=CNTR] ‘instead’

The glosses ‘CNTR’ for “contrastive” for =*ko* and ‘FOC’ for “focus” for =*ga* are thus at best only mnemonic. As =*ko* generally has a contrastive flavor to it, this gloss was chosen for it and, as =*ga* is much more general, being very common with adjuncts of all kinds in which no contrast or topic change is involved, we have opted to gloss it as a general focus particle. Neither of these two glosses is exact, although neither is entirely inappropriate. Perhaps further research will be able to shed more light on these and the many other more seldomly occurring pragmatic markers.

APPENDIX

TEXT: THE NINE TOTEMS [AK, 1]

The following is one of the many versions of the story of the origin of the nine Kharia clans and their respective totems. Although these are not the only Kharia family names, they are considered the original clans and are said to go back to the nine brothers mentioned in this story.

Every version of this story is slightly different. For example, the nine sons here are the offspring of King Sembho and Queen Dakay (see below, note 2), whereas in another version in our corpus ([MS, 1,23ff.]), the father's name is Hondo (*hondo*) and the mother's name is Jhariyo (*jhariyo*).

For further versions, see (in chronological order of appearance) Roy & Roy (1937:130ff., especially 137–139), and in Hindi Kullū (1988 [2000]:124f.) and especially Dṛṇḍṇ (1999:66–70), which presents a number of different versions. For further discussion of the clans and their hierarchical status, see Pfeffer (1993).

Speaker—Anil Birendra Kullu (m.), 30, Saldega, Simdega District, Jharkhand

The language of this speaker is virtually identical to what is generally considered Standard Kharia, both in terms of pronunciation as well as morpho-syntax. The few deviations from this “standard” are noted in the footnotes in the text where they occur.

1. *tama iŋ aniŋ=aʔ gotar=aʔ kahni¹ batay=na col=taʔjɔ=ɪŋ.*
 now 1sg 1PL.INCL=GEN clan=GEN story tell=INF go=MID.PROG=1sg
 ‘Now I am going to tell the story of our clan.’

2. *kahani lebu=ki khoŋi=ki=te kayom=ta=ki. u=ghay*
 story person=PL village.section=PL=OBL speak=MID.PRS=PL this=way
ayiʔj.
 QUAL.PRS
 ‘The people tell [this] story in the villages. It goes like this.’

¹ Standard form: *kahani*.

3. *moṅ raja rani aw=ki=kiyar.*
 one king queen QUAL=MID.PST=DU
 ‘There were a king and queen.’
4. *raja=ya? jimi aw=ki sembho oḍo[?] rani=ya? jimi*
 king=GEN name QUAL=MID.PST Sembho and queen=GEN name
aw=ki ḍakay.
 QUAL=MID.PST Dakay
 ‘The king’s name was Sembho and the queen’s name was Dakay.’²
5. *u ḍakay rani=ya? jimi buṅ=ga ikon, marc mahina=te*
 this Dakay queen=GEN name INST=FOC umh March month=OBL
ḍokḍa? gam=te=ki. acha.
 Dokda? say=ACT.PRS=PL good
 ‘Because of this queen’s name, umh, they call the month of March
 “Dokda?” OK.’
6. *la? u sembho ro ḍakay rani=kiyar=a? nāw jhan*
 then this Sembho and Dakay queen=DU=GEN nine CLASS
be?ṭ=ḍom=kiyar³ aw=ki=kiyar. nāw jhan. thomsij be?ṭ=ḍom=ki.
 SON=3POSS=HON QUAL=MID.PST=HON nine CLASS nine SON=3POSS=PL
 ‘And this Sembho and queen Dakay had nine sons (HON). Nine. Nine
 sons.’

² In the present story these two figures are the parents of the nine sons who founded the nine Kharia clans, whereas in another version in our corpus ([MT, 1]) they are the king and queen of Railogarh (= Rohitasgarh?) who were helped by the Kharia, i.e., there they are not Kharia, much less the parents of the founders of the nine Kharia clans.

According to other authors (e.g. Kullū, 1988 [2000: 37ff.]) they were god-like beings, living in swampy areas, in some versions in Pātāl, the underworld, who saved the Kharia from the fire-rain that God had rained down on humanity to destroy it. Further, as Pinnow (1965a: 148, note b, 6) notes, *sembho* / *sembhu* undoubtedly derives from *sambhu*, one of the many names of Lord Shiva. Further research is necessary here. On the other hand, the form *ḍakay* is closely connected to witches, which have played a key role in traditional Kharia society and are still believed in by many people.

³ Although the form <be²ḍ> would be consistent with the phonological system I propose, this morpheme is consistently written <be²ṭ>, as it derives from the Indo-Aryan form *beṭā*.

7. *sou^ʔb se maha=qom=a^ʔ jimi aw=ki semra.*
 all ABL big=3POSS=GEN name QUAL=MID.PST Semra
 ‘The oldest son’s name was Semra.’
8. *ho=kaṛ=a^ʔ tay konon beghma, ho=kaṛ=a^ʔ tay konon*
 that=SG.HUM=GEN ABL small Beghma that=SG.HUM=GEN ABL small
poṛha, pādra, koynar, ginir, amkul, banari ro
 Porha(= the porha.fruit) Pandra Koynar Ginir Amkul Banari and
cho^ʔdqda=^ʔ be^ʔt=qom=a^ʔ jimi aw=ki baroya.
 small=FOC son=3POSS=GEN name QUAL=MID.PST Baroya
 ‘Younger than him [was] Beghma, younger than him Porha, [and then in decreasing order] Pandra, Koynar, Ginir, Amkul, Banari and the name of the youngest was Baroya.’
9. *moṇ dinu aba=qom sou^ʔb=te remakh=o^ʔ ro gam=o^ʔ*
 one day father=3POSS all=OBL call=ACT.PST and say=ACT.PST
no “dhāy[=e]=pe. u=ghay jinis=te tar ol=e=pe
 Cmpl hurry=ACT.IRR=2PL this=way animal=OBL kill bring=ACT.IRR=2PL
kinir tay” no ho=kaṛ leṅ=ga leṅga col=ta.
 forest ABL and that=SG.HUM fly=FOC REP go=MID.PRS
 ‘One day their father called them all and said “Hurry, kill and bring back this kind of animal from the forest”, and he hurries off (= goes flying).’
10. *aba=qom=kiyar um=kiyar batay=o^ʔ no i jinis*
 father=3POSS=HON NEG=HON tell=ACT.PST Cmpl what animal
heke, i jhāut heke lekin, muda, qaṅ goṭh=o^ʔ
 QUAL.PRS what animal QUAL.PRS but but send C:TEL=ACT.PST
sou^ʔb=te=ga.⁵
 all=OBL=FOC

⁴ Standard form is *choṛka*, a loan word from Sadri. Note also the presence of the non-standard focus marker *=(a)ʔ*, which most speakers will not accept in interviews and generally claim to be unfamiliar with, although its use is by no means uncommon.

⁵ Note that this speaker, like most others, is consciously trying to avoid using common Sadri loan words in his text, “correcting” himself on many occasions, replacing for example *jinis* with *jhāwt* (although in fact both ultimately derive from Sadri), *lekin* with *muda*, etc.

‘Their father didn’t tell them which animal it is, which animal it is, but he sent them all off.’

11. *ro beʔt=dom=ki=jo umay juŋ=oʔ no i jhãut heke*
and son=3POSS=PL=ADD NEG.3PL ask=ACT.PST CMPL what animal QUAL.PRS

lekin kaʔ kom=ki dhoʔ=ke muʔ goʔd=ki=may.
but bow arrow=PL grab=SEQ emerge C:TEL=MID.PST=3PL
‘And his sons also didn’t ask which animal it is, but, having taken their bows and arrows, they went off.’

12. *ro co=naʔ bhere u goʔjhunʔ=te socay=ga col=ki=ki*
and go=INF time this path=OBL think=FOC go=MID.PST=PL

no “i jhãut heke hoy”?
CMPL what animal QUAL.PRS INFER

‘And when they were walking, they walked along this path thinking “What animal could it be?”’

13. *porha gam=oʔ no “sahayda, ikon, khajar, hiran=te,”*
Porha say=ACT.PST CMPL maybe umh deer deer=OBL

tar=na gam=siʔ hoy”.
kill=INF say=PERF INFER

‘Porha said “Maybe, umh, he said to kill a deer, a deer.”’

14. *khariya buŋ “khajar” gam=te=ki.*
Kharia INST “khajar” say=ACT.PRS=PL

‘In Kharia they say *khajar* [for ‘deer’].’

15. *laʔ souʔb=ga bhai=ki ho khajar tar=na ghaʔ juda*
then all=FOC brother=PL that deer kill=INF PURP separate

juda muʔ=ki=may, kinir=te.
REP emerge=MID.PST=3PL forest=OBL

‘Then all the brothers set out separately to kill that deer, into the forest.’

⁶ The lexemes *col* ‘go’ and *del* ‘come’ have the forms *co* and *de*, respectively, before enclitics beginning with /n/.

⁷ Cf. Sadri *harin*.

16. *ro behar=jo umay kuy=o?*, *tay aba=qom ek, mo?*
 and who=ADD NEG.3PL find=ACT.PST then father=3POSS one one
maha daru=ya? tuta=te aw=ki, sou²b=te, sou²b
 big tree=GEN bottom=OBL QUAL=MID.PST all=OBL all
be²t=qom=ki=te, oqo? meson remakh=o? ro gam=o? no:
 son=3POSS=PL=OBL again once call=ACT.PST and say=ACT.PST Cmpl
 ‘And no one found [a deer], then their father was at the bottom of a
 big tree, he called all the sons once again and said:’
17. *“dhāy[=e]=pe”, muda lam=na lamna sou²b=te iku?*
 hurry=ACT.IRR=2PL but search=INF REP all=OBL very
jughay qa? piyas la?²=ki.
 much water thirst EMOT=MID.PST
 ‘“Hurry” but searching and searching, all became thirsty for water.’
18. *tay ho jhāut=te tar=na melay=ke ho=ki qa?²=ya?*
 then that animal=OBL kill=INF stop=SEQ that=PL water=GEN
khoj bu? col=ki=may.
 search INST go=MID.PST=3PL
 ‘Then they gave up killing that animal and went searching for water.’
19. *juda juda, qisa[?]* *ti²j=ga.*
 separate REP far side=FOC
 ‘Separately, far away.’
20. *tay, ikon, ho=ki apan mo?jhi=te kayom=ki=may no*
 then umh that=PL REFL middle=OBL speak=MID.PST=3PL Cmpl
“a=ka? se? qa? kuy=e, ho=ka? u
 CREL=SG.HUM first water find=ACT.IRR that=SG.HUM this
daru=te=ga yo=ga de=na.⁸
 tree=OBL=FOC see=FOC come=MID.IRR
 ‘Then, umh, they said amongst themselves “He who first finds water,
 he should come to this tree, looking [for the others].”’

⁸ Cf. footnote 6 above.

21. *ro batay[=e] no a=boʔ=te qaʔ ayiʔj.*
 and tell=ACT.IRR CMPL Q=place=OBL water QUAL.PRS
 ‘And he should say where the water is.’
22. *souʔb se maha beʔt=dom, simra, moʔn maʔa=te qaʔ*
 all ABL big son=3POSS Semra one cave=OBL water
kuy=oʔ. aqi uʃh=oʔ.
 find=ACT.PST ANAPH drink=ACT.PST
 ‘The eldest son, Semra, found water in a cave. He drank it.’
23. *uq=na loʔqho aqi ho maha daru tuta=te=ga del=ki.*
 drink=INF after ANAPH that big tree bottom=OBL=FOC come=MID.PST
 ‘After drinking, he came to the bottom of that big tree.’
24. *del=ke yo=te laʔ ho=te=ko behar=jo*
 come=SEQ see=ACT.PRS then that=OBL(= ‘there’)=CNTR who=ADD
umboʔiʔj=may.
 NEG.QUAL.PRS=3PL
 ‘Having come, he sees that there is no one there.’
25. *souʔb apan apan muʔ=siʔ=may, qaʔ lam=na.*
 all REFL REP emerge=PERF=3PL water search=INF
 ‘All have set out on their own, to look for water.’
26. *umay eʃ=siʔ.*
 NEG.3PL return=PERF
 ‘They haven’t returned.’
27. *ghaq aqi=te baʔa kharab laʔ=ki.*
 therefore ANAPH=OBL big bad EMOT=MID.PST
 ‘Therefore, he became very sad.’
28. *tay moʔn moʔn bhai=ki=yaʔ jimi qhoʔq=ta qhoʔqta ho=kaʔ*
 then one REP brother=PL=GEN name take=CVB REP that=SG.HUM
ʃhisa=na maʔre=yoʔ.
 call.out=INF start=ACT.PST
 ‘Then, taking the names of the brothers one by one, he began to call out [to them].’

29. *ʃhisa=na bhere... ʃhisa=yoʔ jorse=ga*
 call.out=INF time call.out=ACT.PST loudly=FOC
 ‘When he called . . . , he called out loudly.’
30. *laʔ aɟi=yaʔ tay konon bhai, beghma, ho=kaʔ*
 then ANAPH=GEN ABL small brother Beghma that=SG.HUM
onɟor=oʔ. ro ɟel=ki ho daru boʔ=te.
 hear=ACT.PST and come=MID.PST that tree place=OBL
 ‘Then his younger brother, Beghma, he, heard him. And he, came to the tree.’
31. *tay konon bhai=te juŋ=oʔ no “beghma, am ɟaʔ*
 then small brother=OBL ask=ACT.PST CMPL Beghma 2SG water
uʃh=oʔbʔ”
 drink=ACT.PST.2SG
 ‘Then he asked his younger brother “Beghma, did you drink water?”’
32. *la[ʔ] beghma gam=o[ʔ] “iŋ=ko ɟaʔ=ga um=iŋ kui=siʔ.*
 then Beghma say=ACT.PST 1SG=CNTR water=FOC NEG=1SG find=PERF
laʔ i=ghay uɟ[=e]=iŋ tay?”
 then what=way drink=ACT.IRR=1SG then
 ‘Then Beghma said “I have not found water. Then how should I drink?”’
33. *simra batay=oʔ “dhãy[=e] hẽʔɟõ, ikon, hẽʔɟõ u rochoʔb*
 Semra tell=ACT.PST hurry=ACT.IRR here umh here this side
ɟaʔ ayiʔj.”
 water QUAL.PRS
 ‘Semra said “Hurry! Here, umh, over here there is water.”’
34. *tay beghma col=ki ro ho maʔa boʔ=te ɟam=ki.*
 then Beghma go=MID.PST and that cave place=OBL arrive=MID.PST
 ‘Then . . . Beghma went and he arrived at the cave.’
35. *ɟaʔ=te yo=yoʔ lekin, muda ɟaʔ boʔ=te moŋ*
 water=OBL see=ACT.PST but but water place=OBL one

maha-ɖu[ʔ] kiʔoʔ=te yo=yoʔ

big-APPROX tiger=OBL see=ACT.PST

‘He saw the water but at the water he saw a kind of biggish tiger.’

36. *kiʔoʔ=te, ro botoŋ goʔɖ=ki ro botoŋ=ke eŋ*
tiger=OBL and fear C:TEL= MID.PST and fear=SEQ return

ɖel=ki. enem ɖa[ʔ] uɖ=ga.
come=MID.PST without water drink=FOC

‘A tiger, and he became afraid, and, having become afraid, he returned. Without drinking water.’

37. *ho=ghay=ga konon bhai=te ʃhisa=yoʔ maha beʔt=ɖom,*
that=way=FOC small brother=OBL call.out=ACT.PST big son=3POSS

ə, maha bhai=ɖom, simra, ro poŋha onɖor=oʔ, ɖel=ki
umh big brother=3POSS Semra and Porha hear=ACT.PST come=MID.PST

ho daru boʔ=te.
that tree place=OBL

‘In just this way the elder son, uh, the elder brother, Semra, called the younger brother, and Porha heard him, [and Porha] came to that tree.’

38. *tay aɖi=te=jo batay=oʔ no “dhāy[=e] am, han maŋa*
then ANAPH=OBL=ADD tell=ACT.PST CMPL hurry=A.IRR 2SG that cave

boʔ=te ɖa[ʔ] [a]yiʔj. uɖ ɖe=na=m.”
place=OBL water QUAL.PRS drink come=MID.IRR=2SG

‘Then he said to him, too, “Hurry, at the cave there is water. Drink and come back” ’

39. *laʔ poŋha=jo col=ki ro yo=te ho maŋa boʔ=te*
then Porha=ADD go=MID.PST and see=ACT.PRS that cave place=OBL

ɖam=ki ɖaʔ=te yo=yoʔ.
arrive=MID.PST water=OBL see=ACT.PST

‘“Then Porha also went and looks, he arrived at the cave [and] saw the water.’

40. *lekin moŋ kerketta kontheʔɖ yo=yoʔ.*
but one kerketta bird see=ACT.PST

‘But he saw a *kerketta* bird.’

41. *ho konʰe²d=te yo=yo? ro botoŋ go²d=ki ho=kaɾ.*
 that bird=OBL see=ACT.PST and fear C:TEL=MID.PST that=SG.HUM
 ‘He saw that bird and became afraid.’
42. *ro botoŋ=ke enem ɖa[ʔ] uɖ=ga eŋ ɖel=ki.*
 and fear=SEQ without water drink=FOC return come=MID.PST
 ‘And, being afraid, he came back without drinking water.’
43. *ho=ghay=ga pãdra=jo ho=te maɾa bo²=te*
 that=way=FOC Pandra=ADD that=OBL(= ‘there’) cave place=OBL
ɖam=ki yo=yo? moŋ kulu ayi²j ro enem ɖa[ʔ]
 arrive=MID.PST see=ACT.PST one turtle QUAL.PRS and without water
uɖ=ga eŋ go²d=ki ho=kaɾ=jo.
 drink=FOC return C:TEL=MID.PST that=SG.HUM=ADD
 ‘In this way, Pandra also arrived there at the cave, [and] saw that there is a turtle there and, without drinking water, he also returned.’
44. *ho=kaɾ=a² lo²ɖho koynar col=ki.*
 that=SG.HUM=GEN after Koynar go=MID.PST
 ‘After him, Koynar went.’
45. *koynar=jo moŋ ɖuŋɖuŋ kaɖoŋ=te yo=yo? ro enem*
 Koynar=ADD one eel fish=OBL see=ACT.PST and without
ɖa[ʔ] uɖ=ga botoŋ buŋ eŋ ɖel=ki.
 water drink=FOC fear INST return come=MID.PST
 ‘And Koynar saw an eel (= “eel fish”) and, without drinking water, he returned because of (= “through”) fear.’
46. *ho=kaɾ=a² tay konon ginir col=ki.*
 that=SG.HUM=GEN ABL small Ginir go=MID.PST
 ‘His younger [brother] Ginir went.’
47. *ginir yo=yo? biluŋ=te ro ho=kaɾ=jo botoŋ buŋ*
 Ginir see=ACT.PST salt=OBL and that=SG.HUM=ADD fear INST
eŋ=ki.
 return=MID.PST
 ‘Ginir saw salt and he also returned out of fear.’

48. *ro amkul tetetohoj konthe'd yo=yo?*
 and Amkul tetetohoj bird see=ACT.PST
 'And Amkul saw a tetetohoj bird.'
49. *banari ho=ghay=ga ba? yo=yo?*
 Banari that=way=FOC unhusked.rice see=ACT.PST
 'In just that way Banari saw rice.'
50. *baruya yo=yo? to?po, yo=yo? to?po konthe'd=te.*
 Baroya see=ACT.PST toppo.bird see=ACT.PST toppo bird=OBL
 'Baroya saw a toppo, he saw a toppo bird.'
51. *ro sou'b bhai=ga ho=ghay=ga enem qa[?] uq=ga*
 and all brother=FOC that=way=FOC without water drink=FOC
erj del=ki=may.
 return come=MID.PST=3PL
 'And all brothers thus returned, without drinking water.'
52. *tay aba=qom gam=o[?] no "acha, dhãy[=e]=pe, ab*
 then father=3POSS say=ACT.PST CMPL good hurry=ACT.IRR=2PL now
enem qa[?] uq=ga lam=e=pe ho jinis=te
 without water drink=FOC search=ACT.IRR=2PL that animal=OBL
ro tar=e=pe."
 and kill=ACT.IRR=2PL
 'Then their father said "Good, hurry, now, without drinking water, search for that animal and kill it."'
53. *tay lam=na lamna lamna ho=ki ho jinis=te, khajar,*
 then search=INF REP REP that=PL that animal=OBL deer
kuy=o?=ki ro tar=o?=may.
 find=ACT.PST=PL and kill=ACT.PST=3PL
 'Then searching, searching, searching they found that animal, a deer, and killed it.'
54. *ro ho=ka? a? komar=te bay=o?=ki.*
 and that=SG.HUM=GEN meat=OBL make=ACT.PST=PL
 'And they cut off (= "made") his meat.'

55. *ro souʔb=ga apan apan poʔom bay=oʔ=may.*
 and all=FOC REFL REP bundle make=ACT.PST=3PL
 ‘And all made their own bundles.’
56. *qoʔq=na ghaq taʔj qom=ki=may komaj=te ro*
 take=INF PURP distribute REFL=MID.PST=3PL meat=OBL and
poʔom bay=ke souʔb apan apan hoʔ⁹ muʔ=ki=may.
 bundle make=SEQ all REFL REP house emerge=MID.PST=3PL
 ‘To take it [home with them], they distributed the meat amongst themselves and, having made bundles, they all went (= “emerged”) to their own homes.’¹⁰
57. *tay hoʔ=te aba=qom=te souʔb ob-yo=te=ki no*
 then house=OBL father=3POSS=OBL all CAUS-see=ACT.PRS=PL Cmpl
yo=na ghaq del=ki=may no “ele am=bar
 see=INF PURP come=MID.PST=3PL Cmpl 1PL.EXCL 2=2HON
gam=sikh=oʔ=bar i=ghay khajar tar=na ho jinis=te
 say=PERF=ACT.PST=2HON what=way deer kill=INF that animal=OBL
tar=oʔ=le ro u=ga ho jinis=aʔ komaj.”
 kill=ACT.PST=1PL.EXCL and this=FOC that animal=GEN meat
 ‘Then at home they all show their father [the meat] since they came [for him] to see it. [And they said] “We killed that animal, a deer, the way you (HON) said and this [is] that animal’s meat.”’
58. *tay s[o]uʔb=ga apan apan poʔom=te kholay kholay=ke*
 then all=FOC REFL REP bundle=OBL open REP=SEQ
yo=te=may,
 see=ACT.PRS=3PL
 ‘Then all, having opened their own bundles, look,’

⁹ *hoʔ* for ‘house’ is peculiar to the dialect of this speaker. The standard form is *oʔ*.

¹⁰ The repetition of *apan* here would seem to indicate that each brother went to his own home, although as the following line shows, they returned to their common home to show their father what they had caught.

59. *no je khajar tar=sikh=oʔ=may ho=kaʔ=aʔ¹¹ komaŋ=ko*
 and CREL deer kill=PERF=ACT.PST=3PL that=SG.HUM=GEN meat=CNTR
nalage,
 NEG.QUAL.PRS
 ‘and it isn’t the meat of the deer that they had killed,’
60. *lekin ho maʔa boʔ=te ɖa[ʔ] uɖ=na col=sikh=oʔ=may*
 but that cave place=OBL water drink=INF go=PERF=ACT.PST=3PL
 ‘but at the cave [where] they had gone to drink water,’
61. *ho maʔa boʔ=te je je yo=sikh=oʔ=ki*
 that cave place=OBL CREL REP see=PERF=ACT.PST=PL
 ... *ho=kaʔ=aʔ komaŋ=ga ho poʔom=te aw=ki.*
 that=SG.HUM=GEN meat=FOC that bundle=OBL QUAL=MID.PST
 ‘whatever they had seen at that cave, its meat was in that bundle.’
62. *laʔ aba=ɖom gam=oʔ no “beʔa=ŋ=ki,*
 then father=3POSS say=ACT.PST Cmpl son=1SG=PL
 ‘Then their father said “My sons,’
63. *am=pe, iŋ gam=sikh=oʔj ho=ghay=ga am=pe col=ki=pe*
 2=2PL 1SG say=PERF=ACT.PST.1SG that=way=FOC 2=2PL go=MID.PST=2PL
 ‘You went just as (= “just that way”) I had told you’
64. *ro lam=oʔ=pe ro ho ... khajar=te tar=oʔ=pe*
 and search=ACT.PST=2PL and that deer=OBL kill=ACT.PST=2PL
 ‘and you searched and . you killed that deer’
65. *ro ho=kaʔ=aʔ komaŋ ol=oʔ=pe.*
 and that=SG.HUM=GEN meat bring=ACT.PST=2PL
 ‘and you brought his meat.’
66. *u=te=ko je am=pe yo=sikh=oʔ=pe*
 this=OBL(= ‘here’)=CNTR CREL 2=2PL see=PERF=ACT.PST=2PL

¹¹ As noted in section 5.6, although the “singular human” form =kaʔ is generally restricted to human referents, in line with its original meaning ‘person’, it is occasionally used in the speech of some speakers with non-human referents, as here.

ho=kaɾ=aʔ komaŋ ayiʔj,
 that=SG.HUM=GEN meat QUAL.PRS
 ‘But here is the meat of that [animal] which you saw,’

67. *je musa tay am=pe nãw u=naʔ kuʔum bone=taʔj=pe.*
 CREL today ABL 2=2PL nine this=GEN family become=MID.PROG=2PL
 ‘which as of today you nine are becoming part of its family.’

68. *je je yo=sikh=oʔ=pe ho=je[ʔ]=ga am=p=aʔ kuʔum*
 CREL REP see=PERF=ACT.PST=2PL that=SG.NHUM=FOC 2=2PL=GEN family
heke.
 QUAL.PRS
 ‘Whatever you saw, it is your family.’

69. *ro tama am=pe u nãw kuʔum=te=ga sadi biha*
 and now 2=2PL this nine family=OBL=FOC marry marry
kerson=na=pe ro bes-boʔ, baru-boʔ aw=na=pe.”
 marry=MID.IRR=2PL and good-INTENS good-INTENS live=MID.IRR=2PL
 ‘And you will marry in only these nine families and you will be very
 happy (= “live very well”).’

70. *u=je[ʔ]=ga heke khariya gotar=aʔ choʔka kahani.*
 this=SG.NHUM=FOC QUAL.PRS Kharia clan=GEN small story
 ‘And this is the short history of the Kharia clan.’

[The following text was added at a later date, after I enquired what had happened to Semra at the cave where he drank water (see lines 22–23 above).]

71. *simɕa souʔb se maha beʔt=qom souʔb tay sij¹² ho maɕa*
 Semra all ABL big son=3POSS all ABL first that cave
boʔ=te col=ki ro soreŋ=aʔ moʔjhi tay qaʔ-m-ud=na
 place=OBL go=MID.PST and stone=GEN middle ABL water-m-drink=INF
laʔ=ki.
 IPFV=MID.PST

¹² Standard form: *seŋ*.

‘Semra, the eldest brother, went to that cave first of all [the brothers] and he was drinking water from the middle of the stones.’

72. *ho qaʔ=te uʔh=o[ʔ].*
 that water=OBL drink=ACT.PST
 ‘He drank the water.’

73. *ghađ ađi je bhere eŋ=ki r[o] aba=đom=te*
 therefore ANAPH CREL time return=MID.PST and father=3POSS=OBL
souʔb bhai=kiyar=aʔ soriʔ¹³ pođom=te kholay kholay
 all brother=HON=GEN together bundle=OBL open REP
ob-yo=na laʔ=ki=may, se bhere ađi=yaʔ pođom=te
 CAUS-see=INF IPFV=MID.PST=3PL that time ANAPH=GEN bundle=OBL
soreŋ kui=ki.
 stone find=MID.PST

‘Therefore, when he returned, and along with all his brothers they opened the bundles and were showing them to their father, at that time he [unexpectedly]¹⁴ found a stone in his own bundle.’

¹³ Standard form is *sori*, here perhaps with the nonstandard focus marker *=(a)ʔ* ?

¹⁴ The use of the middle voice with *kui* here denotes an unintentional action or event.

BIBLIOGRAPHY

- Abbi, Anvita. 1993. "Language contraction, language shrink, and language conflation: A case study of Kharia." Mrinal Miri (ed.). *Continuity and change in tribal society*. Simla: Indian Institute of Advanced Study. 542–553.
- . 1997. "Languages in contact in Jharkhand." Anvita Abbi (ed.), 131–148.
- (ed.). 1997. *Languages of tribal and indigenous peoples of India. The ethnic space*. Delhi: Motilal Banarsidass.
- Aikhenvald, Alexandra Y. 2002. "Typological parameters for the study of clitics, with special reference to Tariana." Dixon & Aikhenvald (eds.), 42–78.
- Anderson, Gregory D.S. 2001. "A New classification of South Munda: Evidence from comparative verb morphology." *Indian Linguistics* 62: 21–36.
- . 2007. *The Munda verb. Typological perspectives*. Berlin / New York: Mouton de Gruyter. (Trends in Linguistics, Studies and Monographs, 174).
- . (ed.). 2008. *The Munda languages*. London / New York. (Routledge Language Family Series).
- Anderson, Gregory D.S. & Norman H. Zide. 2002. "Issues in Proto-Munda and Proto-Austroasiatic nominal derivation: the bimoraic constraint." Marlys A. Macken (ed.), *Papers from the 10th annual meeting of the Southeast Asian Linguistics Society*. Tempe, AZ: Arizona State University, South East Asian Studies Program. (Monograph Series Press). 55–74.
- Anderson, Stephen R. 2005. *Aspects of the theory of clitics*. Oxford: Oxford University Press. (Oxford Studies in Theoretical Linguistics, 11).
- Andrews, Avery. 1985. "The major functions of the noun phrase." Timothy Shopen (ed.). *Language typology and syntactic description, Volume I: Clause Structure*. Cambridge: Cambridge University Press. 62–154.
- Archer, W.G. (ed.). 1942. *khariyā aloṅ. A Kharia song book*. Ranchi: Catholic Press.
- Ba?, Juliys. 1983. *Khariyā dhvaniśāstra. Kharia phonetics*. Ranchi: Merī Silīm.
- . 2001. *Khariyā dhvani sārācnā kā vajñānik viśleṣaṅ [A scientific analysis of the structure of Kharia sounds]*. Ranchi: Dāso Khariyā.
- Banerjee, Gagan Chandra. 1894. *Introduction to the Khariā Language*. [Reprint: New Delhi / Chandigarh: Bahri Publications, 1982].
- Bashir, Elena. 2003. "Dardic" Cardona & Jain (eds.), 818–894.
- Benveniste, Émile. 1966. "Chapitre XIII. La phrase nominale." *Problèmes de linguistique générale*. Paris: Éditions Gallimard. [Reprint from *Bulletin de la Société de Linguistique de Paris*, XLVI (1950), fasc. I, n° 132].
- Berlin, Brent & Paul Kay. 1969. *Basic color terms. Their universality and evolution*. Berkeley / Los Angeles: University of California Press.
- Bhagat, Sureś. 2001a. "Sāvar." *Bulletin of Jharkhand Tribal Welfare Research Institute, Ranchi*. 39: 18–20.
- . 2001b. "Hil Khariyā" ["Hill Kharia."]. *Bulletin of Jharkhand Tribal Welfare Research Institute, Ranchi*. 39: 21–26.
- Bhattacharya, Sudhibhushan. 1968. *A Bonda dictionary*. Poona: Deccan College Post-Graduate and Research Institute. (Building Centenary and Silver Jubilee Series, 18).
- Biligiri, H.S. 1965. *Kharia: Phonology, grammar and vocabulary*. Poona: Deccan College, Postgraduate and Research Institute. (Building Centenary and Silver Jubilee Series, 3).
- Bisang, Walter. 1998. "Adverbiality: The view from the Far East." Johan van der Auwera & Dónall P. Ó Baoill (eds.), *Adverbial constructions in the languages of*

- Europe. Berlin / New York: Mouton de Gruyter (Empirical Approaches to Language Typology, Eurotyp 20–3). 641–812.
- Bloomfield, Leonard. 1933 [1984]. *Language. With a new foreword by C.F. Hockett.* Chicago / London: The University of Chicago Press.
- Bril, Isabelle. 2005. “Semantic and functional diversification of reciprocal and middle prefixes in New Caledonian and other Austronesian languages.” *Linguistic Typology* 9: 25–76.
- Broschart, Jürgen. 1997. “Why Tongan does it differently: Categorical distinctions in a language without nouns and verbs.” *Linguistic Typology* 1: 123–165.
- Cardona, George & Dhanesh Jain (eds.). 2003. *The Indo-Aryan languages.* London / New York: Routledge (Routledge Language Family Series).
- Catholic Biblical Association of America. 1970. *Saint Joseph edition of the New American Bible. Translated from the original languages with critical use of all the ancient sources.* New York: Catholic Book Publishing Co.
- Croft, William. 1991. *Syntactic categories and grammatical relations. The cognitive organization of information.* Chicago / London: University of Chicago Press.
- . 2000. “Parts of speech as language universals and as language-particular categories.” Petra M. Vogel & Bernard Comrie (eds.), *Approaches to the typology of word classes.* Berlin / New York: Mouton de Gruyter (Empirical Approaches to Language Typology, 23). 65–102.
- . 2005. “Word classes, parts of speech, and syntactic argumentation.” *Linguistic Typology* 9: 431–41.
- Deeney, J., S.J. 1975. *Ho grammar and vocabulary.* Chaibasa: Xavier Ho Publications, St. Xavier’s High School.
- . 1978. *Ho-English dictionary.* Chaibasa: Xavier Ho Publications, St. Xavier’s High School.
- Delancey, Scott. 1997. “Mirativity: The grammatical marking of unexpected information.” *Linguistic Typology* 1: 33–52.
- Di Sciullo, Anna & Edwin Williams. 1987. *On the definition of word.* Cambridge, MA / London: MIT.
- Dixon, R.M.W. 1994. *Ergativity.* Cambridge: Cambridge University Press (Cambridge Studies in Linguistics, 69).
- Dixon, R.M.W. & Alexandra Y. Aikhenvald. 2002a. “Word: a typological framework.” Dixon & Aikhenvald (eds.), 1–41.
- (eds.). 2002b. *Word. A cross-linguistic typology.* Cambridge: Cambridge University Press.
- Donegan, Patricia & David Stampe. 2004. “Rhythm and the synthetic drift of Munda.” Rajendra Singh (editor-in-chief), *The Yearbook of South Asian Languages and Linguistics, 2004.* Berlin / New York: Mouton de Gruyter. 3–36.
- Dunḍuḍ, Phā. Jovākīm. 1999. *Khariyā jīvan aur paramparāē [Kharia life and traditions].* Ranchi: Ranchi Society of Jesus.
- Dunḍuḍ, Mathiyas. 1986. *Hindī aur Khariyā (tulnātmak aur viśeṣaṇātmak adhyāyan) [Hindi and Khariya. A comparative and analytical study.]* Ranchi: Satya Bhāratī.
- Ebert, Karen H. 1993. “Kiranti subordination in the South Asian areal context.” Karen H. Ebert (ed.). *Studies in clause linkage. Papers from the first Köln-Zürich Workshop.* Zürich: ASAS-Verlag. 83–110.
- . 1998. *Finite forms in non-finite function: The Tibeto-Burman-Dravidian-Munda contact area.* Zürich: Allgemeine Sprachwissenschaft. (Working Paper, No. 7).
- . 1999. “Nonfinite verbs in Kiranti languages—an areal perspective.” In: Yogendra P. Yadava & Warren W. Glover (eds.). *Topics in Nepalese linguistics.* Kathmandu: Royal Nepal Academy. 371–400.
- Ebert, Karen H. & Lukas Neukom. 2000. *Towards a typology of relative clauses in South Asia.* Zürich: Allgemeine Sprachwissenschaft (Working Paper, No. 9).
- Emeneau, M.B. 1956. “India as a linguistic area.” *Language* 32/1: 3–16.

- Evans, Nicholas & Toshiki Osada. 2005a. "Mundari: the myth of a language without word classes." *Linguistic Typology* 9: 351–390.
- . 2005b. "Mundari and argumentation in word-class analysis." *Linguistic Typology* 9: 442–457.
- Falk, Yehuda N. 2001. *Lexical-Functional Grammar. An introduction to parallel constraint-based syntax*. Stanford, California: CSLI.
- Fernandez, Frank. 1967. *A grammatical sketch of Remo: A Munda language*. Unpublished Ph.D. dissertation, Chapel Hill, North Carolina.
- Filbek, David. 1976. "Toward a grammar of relative clauses in T'in." Philip N. Jenner, Laurence C. Thompson & Stanley Starosta (eds.). *Austroasiatic Studies*. Honolulu: University Press of Hawaii. 1: 285–307.
- Floor, H., S.J., V. Gheysens, S.J. & G. Druart, S.J. 1934. *Tea Districts Labour Association. Dictionary of the Kharia language*. Calcutta: Begg, Dunlop & Co., Ltd.
- Foley, William A. & Robert D. Van Valin, Jr. 1985. "Information packaging in the clause." Timothy Shopen (ed.). *Language typology and syntactic description, I: Clause Structure*. Cambridge: Cambridge University Press. 282–364.
- Gair, James W. 2003. "Sinhala" Cardona & Jain (eds.), 766–817.
- Genetti, Carol. 1991. "From postposition to subordinator in Newari." Elizabeth Closs Traugott & Bernd Heine (eds.). *Approaches to grammaticalization. Volume II: Types of grammatical markers*. Amsterdam / Philadelphia: John Benjamins. (Typological Studies in Language, 19). 227–255.
- Gonda, J. 1979. *The medium in the Rgveda*. Leiden: E.J. Brill (Orientalia Rheno-Traiectina, 24).
- Grierson, G.A. 1906. *Linguistic Survey of India. Vol. IV. Muṇḍā and Dravidian languages*. [Reprint: Delhi: Low Price Publications, 1994].
- Grimes, Barbara F. (ed.). 1988. *Ethnologue*. Dallas: Summer Institute of Linguistics.
- Haegeman, Liliane. 1991. *Introduction to Government and Binding Theory. Second edition*. Malden, MA: Blackwell. (Blackwell Textbooks in Linguistics, 1).
- Hengeveld, Kees. 1992. "Parts of speech." Michael Fortescue, Peter Harder & Lars Kristoffersen (eds.). *Layered structure and reference in a functional perspective. Papers from the functional grammar conference in Copenhagen 1990*. Amsterdam / Philadelphia: Benjamins. (Pragmatics & Beyond, New Series, 23). 29–55.
- Hengeveld, Kees & Jan Rijkhoff. 2005. "Mundari as a flexible language." *Linguistic Typology* 9: 406–431.
- Hoffmann, Revd. J., S.J. 1903. *Mundari grammar*. Calcutta: Bengal Secretariat Press.
- . 1905 / 1909. *Mundari grammar and exercises. Parts I & II*. [Reprint: New Delhi: Gyan Publishing House, 2001].
- Hopper, Paul J. & Sandra A. Thompson. 1980. "Transitivity in grammar and discourse." *Language* 56/1: 251–299.
- Ishtiaq, M. 1999. *Language shifts among the scheduled tribes in India. A geographical study*. Delhi: Motilal Banarsidass. (MLBD Series in Linguistics, 13).
- Jordan-Horstmann, Monika. 1969. *Sadani. A Bhojpuri dialect spoken in Chotanagpur*. Wiesbaden: Harrassowitz (Indologie Berolinensis, 1).
- Kaufmann, Ingrid. 2004. *Medium und Reflexiv: Eine Studie zur Verbsemantik*. Tübingen: Niemeyer.
- Keenan, Edward L. 1976. "Towards a universal definition of 'subject'." Charles N. Li (ed.). *Subject and topic*. New York: Academic Press. 305–333.
- . 1985. "Relative clauses." Timothy Shopen (ed.). *Language typology and syntactic description. II: Complex constructions*. Cambridge: Cambridge University Press. 141–170.
- Kemmer, Suzanne. 1993. *The middle voice*. Amsterdam / Philadelphia: Benjamins (Typological Studies in Language, 23).

- Kerkeṭṛā, Khrist Pyāri. 1990. *Jujhair dāṛ* (*Khaṛiyā nāṭak*) [*The Battle Field (A Kharia Drama)*]. Ranchi: Tribal Language Academy, Government of Bihar.
- Klaiman, Miriam H. 1991. *Grammatical voice*. Cambridge: Cambridge University Press (Cambridge Studies in Linguistics, 59).
- Koptjevskaja-Tamm, M. 1994. "Finiteness." R.E. Asher (editor-in-chief), *The encyclopedia of language and linguistics*. Vol. 3. Oxford: Pergamon Press. 1244–1248.
- Kruspe, Nicole. 2004. *A grammar of Semelai*. Cambridge: Cambridge University Press. (Cambridge Grammatical Descriptions).
- Kullū, Anil Virendra. 2007. *sorej khoṛi*. [*Stone Village*]. Ranchi: Satya Bhārati.
- Kullū, Paulus. 1981. *Khaṛiyā vyākaraṇ evā sāksipta śabdakoś* [*Kharia grammar and abridged dictionary*]. Ranchi: Religious Literature Society.
- 1992. *Tonne kol-āṅgre* [*The New Testament in Kharia*]. Ranchi: Catholic Press.
- 1988. [Reprint 2000]. *Khaṛiyā dharam aur sāskṛti kā viśleṣaṇ* [*An analysis of Kharia religion and culture*]. Ranchi: Satya Bharati.
- Lambrecht, Knud. 1994. *Information structure and sentence form. Topic, focus and the mental representations of discourse referents*. Cambridge: Cambridge University Press. (Cambridge Language Surveys).
- Lazard, Gilbert. 1999. "Mirativity, evidentiality, mediativity, or other?" *Linguistic Typology* 3: 91–109.
- Lehmann, Christian. 1995. *Thoughts on grammaticalization*. München / Newcastle: Lincom Europa. (LINCOM Studies in Theoretical Linguistics, 1).
- Levin, Beth & Malka Rappaport Hovav. 1995. *Unaccusativity. At the syntax-lexical semantics interface*. Cambridge, Massachusetts / London, England: MIT.
- Lewis, M. Paul (ed.), 2009. *Ethnologue: Languages of the world*, Sixteenth edition. Dallas, Texas: SIL International. Online version: <http://www.ethnologue.com/>.
- Lichtenberk, Frantisek. 1985. "Multiple uses of reciprocal constructions." *Australian Journal of Linguistics* 5: 19–41.
- Maas, Utz. 2004. "Finite" and "nonfinite" from a typological perspective. *Linguistics* 42/2: 359–385.
- 2007. "Mirative im Maltesischen—ein Fall von Grammati(kali)sierung?" *Linguistische Berichte*, 210: 129–146.
- Mahapatra, B.P. 1976. "Comparative notes on Juang and Kharia finite verbs." Philip N. Jenner, Laurence C. Thompson, and Stanley Starosta (eds.). *Austroasiatic studies*. Honolulu: University of Hawaii (Oceanic Linguistics, Special Publication, No. 13), 801–814.
- Malhotra, Veena. 1982. *The structure of Kharia: A study of linguistic typology and language change*. Unpublished Ph.D. dissertation. New Delhi: Jawaharlal Nehru University.
- Masica, Colin P. 1976. *Defining a linguistic area. South Asia*. Chicago / London: University of Chicago Press.
- Matson, Dan Mitchell. 1964. *A grammatical sketch of Juang, a Munda language*. Unpublished Ph.D. dissertation. University of Wisconsin.
- Michailovsky, Boyd. 1996. "L'Inférentiel du népali." Zlatka Guentchéva (éd.): *L'Énonciation médiatisée*. Louvain—Paris: Éditions Peeters (Bibliothèque de l'Information Grammaticale). 109–123.
- Montaut, Annie. 2004. *A grammar of Hindi*. München: Lincom Europa. (LINCOM Studies in Indo-European Linguistics, 2).
- Moral, Dipankar. 1997. "North-East India as a linguistic area." *Mon-Khmer Studies* 27: 43–53.
- Narasimhan, Bhuvana. 2003. "Motion events and the lexicon: a case study of Hindi." *Lingua* 113: 123–160.
- Neukom, Lukas. 1999. "Phonological typology of Northeast India." *Linguistics of the Tibeto-Burman Area* 22/2: 121–147.

- . 2001. *A grammar of Santali*. München: Lincom Europa (Languages of the World/Materials, 323).
- Nichols, Johanna, David A. Peterson & Jonathan Barnes. 2004. "Transitivizing and detransitivizing languages." *Linguistic Typology* 8/2: 149–211.
- Nowrangi, P.S. 1956. *A simple Sadānī grammar*. Ranchi: D.S.S. Book Depot.
- Osada, Toshiki. 1992. *A reference grammar of Mundari*. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.
- Palmer, F.R. 1994. *Grammatical roles and grammatical relations*. Cambridge: Cambridge University Press.
- Patnaik, Manideepa. 2008. "Juang." In: Gregory D.S. Anderson (ed.), 508–556.
- Peterson, John. 1998. *Grammatical relations in Pāli and the emergence of ergativity in Indo-Aryan*. München / Newcastle: Lincom Europa. (LINCOS Studies in Indo-European Linguistics, 1).
- . 2000. "Evidentials, inferentials and mirativity in Nepali." Balthasar Bickel (ed.). *Person and evidence in Himalayan languages*. Special Edition of the *Linguistics of the Tibeto-Burman Area*. 23/2: 13–37.
- . 2002a. "Morphological and semantic finiteness in Kharia: A first look." Michael Bommes, Doris Tophinke & Christina Noack (eds.). *Festschrift für Utz Maas*. Wiesbaden: Westdeutscher Verlag.
- . 2002b. "The Nepali converbs: A holistic approach." Rajendra Singh (chief editor), *The Yearbook of South Asian Languages and Linguistics*. 2002. New Delhi / Thousand Oaks / London: Sage Publications. 93–133.
- . 2005. "There's a grain of truth in every 'myth', or, Why the discussion of lexical classes in Mundari isn't quite over yet." *Linguistic Typology* 9: 391–405.
- . 2006. *Kharia. A South Munda Language*. Unpublished Habilitationsschrift, Universität Osnabrück.
- . 2007. "Languages without nouns and verbs? An alternative to lexical classes in Kharia." Colin Masicca (ed.). *Old and new perspectives on South Asian languages. Grammar and semantics. Papers growing out of the Fifth International Conference on South Asian Linguistics (ICOSAL-5), held at Moscow, Russia in July 2003*. Delhi: Motilal Banarsidass. 274–303.
- . 2008. "Kharia." Gregory D.S. Anderson (ed.), 434–507.
- . 2009. A Kharia-English Lexicon." *Himalayan Linguistics Archive* 5. <http://www.linguistics.ucsb.edu/HimalayanLinguistics/>
- . in press, a. "Languages without nouns and verbs. A formal account." Jan Rijkhoff & Eva van Lier (eds.), *Flexible word classes: a typological study of underspecified parts-of-speech*.
- . in press, b. "The "totality" morpheme *-may* in Kharia." To appear in Ram Dayal Munda (ed.): *Felicitation volume for Norman Zide*.
- . forthcoming, a. "Language contact in Jharkhand. Linguistic convergence between Munda and Indo-Aryan in eastern-central India."
- . forthcoming, b. "Active and middle in Kharia (South Munda): A prototypical approach."
- Peterson, John & Utz Maas. 2009. "Reduplication in Kharia: the masdar as a phonologically motivated category." *Morphology*, 19:207–237.
- Pfeffer, Georg. 1993. "Kharia totemism." *Journal of the Indian Anthropological Society*. 28: 221–227.
- Pinnow, Heinz-Jürgen. 1959. *Versuch einer historischen Lautlehre der Kharia-Sprache*. Wiesbaden: Otto Harrassowitz.
- . 1960. "Über den Ursprung der voneinander abweichenden Strukturen der Munda- und Khmer-Nikobar-Sprachen." *Indo-Iranian Journal* 4: 81–103.
- . 1963. "The position of the Munda languages within the Austroasiatic language family." H.L. Shorto (ed.), *Comparison in Southeast Asia and the Pacific*. London: SOAS. 140–152.

- . 1965a. *Kharia-Texte (Prosa und Poesie)*. Wiesbaden: Otto Harrassowitz.
- . 1965b. ““Der Wertiger” und andere Geschichten in Kharia, aufgezeichnet von Junas Kullu, übersetzt und mitgeteilt unter Mitwirkung von Paulus Kerketta.” *Indo-Iranian Journal* 9/1: 32–68.
- . 1966. “A comparative study of the verb in the Munda languages.” Norman H. Zide (ed.). *Studies in comparative Austroasiatic linguistics*. The Hague: Mouton & Co. (Indo-Iranian Monographs, V).
- Pustejovsky, James. 1998. *The generative lexicon*. Cambridge, MA / London: MIT.
- Rehberg, Kerstin. 2003. *Phonologie des Kharia—Prosodische Strukturen und segmentales Inventar*. Unpublished M.A. thesis. Osnabrück: Universität Osnabrück.
- Risley, H.H. 1891. *The tribes and castes of Bengal. Ethnographic glossary*. Calcutta: Bengal Secretariat Press [Reprint: Calcutta: P. Mukherjee, 1998].
- Roy, Sarat Chandra & Ramesh Chandra Roy. 1937. *The Khārīās*. Ranchi: Man in India.
- Sadock, Jerrold M. 1991. *Autolexical syntax. A theory of parallel grammatical representations*. Chicago / London: University of Chicago Press. (Studies in Contemporary Linguistics).
- Sāhu, R.P. 1979/80. *Khāṛiyā vyākaraṇ [Kharia grammar]*. Ranchi: Kharia Department, Gossaner College.
- Sasse, Hans-Jürgen. 1978. “Subjekt und Ergativ: Zur pragmatischen Grundlage primärer grammatischer Relationen.” *Folia Linguistica* XII: 3–4. 219–252.
- . 1993. “Das Nomen—eine universale Kategorie?” *Sprachtypologie und Universalienforschung (STUF)* 46/3: 187–221.
- Schultze-Berndt, Eva & Nikolaus P. Himmelmann. 2004. “Depictive secondary predicates in cross-linguistic perspective.” *Linguistic Typology* 8: 59–131.
- Simpson, Jane. 1991. *Warlpiri morpho-syntax. A lexicalist approach*. Dordrecht: Kluwer Academic Publishers. (Studies in Natural Language and Linguistic Theory, 23).
- Sinha, Dikshit. 1984. *The Hill Kharia of Purulia. A study on the impact of poverty on a hunting and gathering tribe*. Calcutta: Anthropological Survey of India, Government of India. (Memoir, Anthropological Survey of India, 59).
- Talmy, Leonard. 1985. “Lexicalization patterns: semantic structure in lexical forms.” Timothy Shopen (ed.), *Language typology and syntactic description. Volume III: Grammatical categories and the lexicon*. Cambridge: Cambridge University Press.
- Tea Districts Labour Association. 1929. *Language hand-book Kharia. Printed for private circulation only*. Calcutta: Begg, Dunlop & Co.
- Van Valin, Robert D., Jr. 2005. *Exploring the syntax-semantics interface*. Cambridge: Cambridge University Press.
- Van Valin, Robert D., Jr. & Randy J. LaPolla. 1997. *Syntax. Structure, meaning and function*. Cambridge: Cambridge University Press. (Cambridge Textbooks in Linguistics).
- Vidyarthi, L.P. & V.S. Upadhyay. 1980. *The Kharia: Then and now. A comparative study of Hill, Dhelki and Dudh Kharia of the central-eastern region of India*. Delhi: Concept Publishing Company.
- Yadav, Ramawatar. 1996. *A reference grammar of Maithili*. Berlin: Mouton de Gruyter. (Trends in Linguistics, Documentation, 11).
- Zide, Norman H. 1968. “Some comparative notes on Gata? pronouns.” J.C. Heesterman, G.H. Schokker & V.I. Subramoniam (eds.). *Pratidānam. Indian, Iranian and Indo-European studies presented to Franciscus Bernardus Kuiper on his sixtieth birthday*. The Hague / Paris: Mouton. 348–358.
- . 1969. “Munda and Non-Munda Austroasiatic languages.” In: T.A. Sebeok (ed.). *Current Trends in Linguistics. Vol. 5: Linguistics in South Asia*. The Hague: Mouton. 411–430.

- . 1985. "Notes on some participant roles in some Munda languages." *Proceedings of the Conference on Participant Roles: South Asia and Adjacent Areas*. Bloomington, Indiana: Indiana University Linguistics Club. 92–103.
- Zide, N.H. & D.L. Stampe. 1968. "The place of Kharia-Juang in the Munda family." Bhadriraju Krishnamurti (ed.). *Studies in Indian linguistics: Professor M.B. Emeneau volume*. Poona: Centres of Advanced Study in Linguistics, Deccan College and Annamalai University. 370–377.

GENERAL INDEX

- ablative / ablativity xxi, 145, 152, 202, 390
- abstract (concepts) 81, 102, 278
- abu* xiii, 249, 258, 334, 338–339, 429
- accent 35, 36, 44
- acceptability x, 112–116, 117
- action(al) 69, 71, 73, 80, 85, 91, 102, 110, 131, 177, 201, 209, 216, 219, 230, 235, 238, 243, 247, 250, 256, 261, 265, 267, 268, 270, 272, 273, 274, 275, 276, 277, 278, 279, 281, 284, 285, 286, 287, 288, 290, 291, 292, 293, 299, 300, 301, 302, 304, 305, 306, 307, 308, 309, 311, 312, 315, 325, 326, 334, 343, 344, 345, 347, 359, 397, 413, 452
- active See *basic voice*
- activity 13, 78, 85, 97, 286, 305
- additive xxi, 53, 137, 179, 182, 395, 434
- adhortative 243, 245, 246, 247
- aḍi* / *aḍ* 11, 54, 114, 138, 152, 160, 162, 166, 169, 171, 172, 173, 174, 175, 184, 187, 202, 217, 258, 308, 312, 313, 320, 322, 330, 342, 350, 354, 390, 391, 407, 409, 412, 422, 444, 445, 446, 452
- adjective See *parts of speech*
- adjectivizer 106
- adjunct xi, 1, 83, 88, 94, 139, 143, 144, 145, 150, 153, 154, 156, 205, 206, 261, 293, 305, 314, 317, 323, 326, 352, 356, 359, 368, 393, 397, 412, 426, 427, 428, 432, 436, 438
- adposition See also *case* xi, 35, 71, 110, 111, 123, 135, 138, 147, 149–158, 212
- postposition 24, 44, 68, 77, 88, 109, 111, 145, 148, 149–158, 178, 192, 202, 204, 220, 235, 236, 324, 366, 367, 389, 390, 403, 455
- preposition 149, 150
- adverb(ial) xi, xiii, 72, 88, 97, 139, 142, 154, 183, 197, 205–207, 308, 326, 331, 368, 389–403, 453
- affix x, xxi, 17, 24, 42, 51, 52, 57, 59, 60, 61, 68–69, 109, 211, 230
- infix x, xxi, 1, 17, 30, 69, 75, 92, 93, 94, 101–105, 180, 192, 210, 230, 231, 233, 234, 235, 414
- prefix 1, 69, 92, 93, 210, 230, 231, 233, 235, 454
- suffix 1, 12, 55, 56, 59, 69, 92, 143, 163, 193, 221, 334
- affricate 29
- agent(ive) 79, 91, 167, 212, 227, 230, 235, 236, 261, 262, 268, 275, 277, 281, 289, 293, 309, 344, 363, 413
- agreement 51, 59, 214, 352, 353
- Aktionsart* xii, xxii, 4, 12, 25, 45, 278, 293
- alienable See *possession*
- allative 146, 154, 157
- allophone 20, 21, 27, 28, 29, 30
- alternative question xiii, 368, 373
- ambulative xii, xxi, 306
- anaphoric xxi, 11, 60, 111, 171–173, 175, 185, 186, 431
- Andaman (Islands) 1
- Andhra Pradesh 6
- animate / animacy (non- or inanimate / non- or inanimacy) xi, 11, 85, 112, 117, 138, 139–142, 175, 176, 177, 186, 191, 199, 222, 223, 224, 263, 280, 285, 287, 290, 346, 347, 413
- anticipatory clitic See *clitic*
- anticipatory telic See *telic*
- anugrah* 21
- apan* 60, 155, 267, 295, 298, 363, 364, 365, 389, 390, 399, 404, 409, 443, 444, 449
- apposition 212, 336, 352, 353, 354, 355, 356, 357
- approximant 29
- approximative xxi, 69, 89, 193–194
- Arabic 96, 318
- argument (of the predicate) xiv, 1, 56, 73, 125, 168, 212, 314, 351, 352–359, 360, 362, 367, 368, 393, 396, 412, 420, 426, 432, 433
- article (definite / indefinite) 18, 69, 142, 334
- Aslian 101

- aspect(ual) xiii, 77, 209, 240, 241,
259, 267, 278, 281, 329, 341,
342–346, 375, 377, 382, 383
- aspirate / aspiration 5, 20, 23, 29, 31,
33, 63, 240
- Assam 1
- associative See *compound*
- attributive 12, 17, 18, 19, 71–118,
120, 162, 175, 179, 184, 200–206,
319, 334, 397, 406–425
- Austro-Asiatic xvii, 7, 8, 11, 15, 92, 101,
118, 425, 453, 455, 456, 458, 459
- autopoesis xi, xxi, 297, 309
- auxiliary xiii, 4, 224, 312, 323, 335,
337, 341, 342–346, 365
- avolitional See *volitional*
- background(ing) 293, 295, 297, 343,
344, 386, 426, 428
- basic TAM-categories xi, 239–258, 259
- basic voice xi, 12, 37, 59, 60, 66, 67,
71, 88, 209, 210, 215, 239, 240, 242,
249, 258–292, 298, 299, 375, 377,
388
- active xii, xxi, 12, 15, 25, 27, 33,
34, 35, 59, 71, 75–91, 95, 96,
100, 201, 209, 214, 215, 221–224,
225, 230, 232, 235, 239, 240, 242,
243, 246, 248, 251, 252, 253, 256,
258–292, 297, 300, 314, 315, 316,
317, 331, 340, 342, 358, 359, 368,
402, 457
- active-only 222, 261, 264, 265, 279,
281–289
- middle xii, xxii, 10, 12, 14, 15, 25,
27, 34, 59, 64, 71, 75–91, 95, 96,
97, 100, 104, 112, 200, 201, 209,
213, 214, 215, 221–224, 225, 230,
239, 240, 242, 243, 246, 249, 251,
253, 256, 258–292, 297, 298, 299,
314, 317, 324, 331, 342, 345, 358,
359, 379, 402, 452, 454, 455, 457
- middle-only 222, 261, 264, 265,
276–287
- benefactive xii, xxi, 309, 318, 356,
358, 367
- Bengal 1, 3, 342, 458
- Bengali 5, 6, 190, 377
- Bible 11, 220, 303, 342, 454
- bimoraic (constraint) 92, 93, 453
- Birmitrapur xviii, 1, 16
- bisyllabic x, 35, 36, 40, 42, 44, 51, 62,
63, 88, 93, 94, 97, 98, 99, 100, 102,
103, 121, 132, 201, 202, 231, 233,
234, 266, 267, 317
- body posture 289, 290, 291, 292
- borrowed 51, 87, 92, 96, 97, 98, 139,
150, 183, 188, 189, 192, 193, 199,
201, 221, 222, 223, 225, 235, 260,
295, 297, 318, 323, 324, 326, 333,
334, 363, 377, 381, 393, 395, 408,
417
- Caddo 401
- cardinal (number) 103, 189, 191, 192,
193, 194, 197, 199
- case See also *adposition*
- direct 4, 88, 135, 143–147, 149,
150, 226, 228, 294, 327, 353, 416,
417, 423
- genitive xxi, 1, 9, 12, 27, 34, 52,
55, 64, 67, 71, 77, 78, 88, 89, 90,
94, 103, 107, 108, 109, 110, 112,
113, 120, 135, 137, 138, 139, 143,
146, 147, 148, 149, 150, 151, 163,
182, 191, 197, 198, 210, 212, 221,
227, 317, 319, 320, 321, 322, 366,
389, 390, 393, 397, 410, 412, 414,
415, 416, 417, 418, 419, 423
- oblique xi, xxii, 1, 55, 56, 66, 97,
135, 142, 143, 144, 145, 146, 147,
148, 150, 151, 153, 154, 156, 178,
183, 206, 220, 226, 227, 228, 235,
236, 294, 295, 321, 327, 331, 348,
362, 364, 367, 368, 397, 420
- Case-syntagma (use of term) 18–19
- cataphoric 171, 172, 185
- causal xiii, 389, 390
- causative xi, xxi, 4, 17, 30, 42, 69,
94, 98, 180, 210, 211, 214, 215, 216,
218, 221, 222, 224, 227, 230–236,
261, 262, 264, 274, 275, 317, 318,
367, 368, 413
- double causative 4, 210, 230,
233–236, 367, 368
- chaining 237, 238
- change of state 77, 201, 253, 256,
278, 279, 286, 299, 302, 304, 305
- Chattisgarh xv, 1, 2, 3
- Chinese 206
- circumnominal xiv, 406, 407,
419–421, 423, 424, 425
- classifier xi, xxi, 137, 138, 142, 187,
188, 189, 193, 194–200
- clause-level 18, 52, 330, 359, 367
- clitics
- anticipatory clitic 62, 67
- clitic word See *word*
- enclitic xxi, 1, 12, 17, 24, 30, 33,
34, 35, 36, 37, 40, 42, 44, 45,

- 51–68, 69, 71, 97, 109, 117, 120,
126, 129, 137, 138, 143, 194, 209,
211, 234, 324, 329, 338, 349, 352,
353, 354, 355, 356, 357, 359, 364,
365, 367, 403, 427, 433–438, 442
mobility *x*, 67, 239
proclitic 24, 44, 53, 408
closed (class) 12, 77, 78, 88, 90, 92,
106, 117, 201
coda 20, 22, 28, 29, 31, 32, 33, 34,
104, 122, 129, 256, 257
cognition active 289
cognition middle 289, 291, 292
collective 161, 162, 193, 198, 237,
239, 290
committative 150, 154
comparative *xi*, 202, 204, 297, 410,
412, 414, 417, 418, 419, 423, 453,
454, 456, 458
complement (of a predicate) 4, 96,
318, 359
complement clause *xiv*, 366, 389, 398,
403–406
complementizer *xxi*, 333, 369, 391,
404, 405
compound *x*, *xxi*, 11, 24, 25, 42, 44,
60, 75, 84, 85, 92, 99, 100, 102, 103,
107, 118–128, 129, 130, 133, 136,
139, 157, 173, 191, 393
associative 118, 119, 120
compound form 92, 102, 103, 121,
124, 130, 191
equational 118
pseudo-compound *x*, 124–128
conative *xii*, *xxi*, 305, 308
concessive *xiv*, 389, 395
conditional *xiv*, 247, 253, 389, 393,
395
configurational 73
conjunct verb *See light verb*
conjunctive participle *See converb*,
sequential
contentive morpheme (use of term) 17
continuative *xii*, *xxi*, 298, 299, 301
contrastive *xxi*, 53, 137, 429, 433–438
converb *xii*, *xxi*, *xxii*, 313, 334, 457
“imperfective” 105, 133, 207,
330–333, 393, 399, 400, 401
sequential 4, 149, 153, 219, 296,
311, 312, 324–330, 361–362, 367,
396, 397, 404
conversion 18
coordination *xiii*, 25, 388
copula *ix*, *xiii*, 12, 19, 25, 34, 72, 76,
97, 210, 336, 349, 374–388
core 314, 325, 330, 358, 359, 360, 367
coreferential 351, 358, 359, 360–361,
362, 367, 420, 425
correlative *xiv*, *xxi*, 186, 392, 395,
406, 407, 408, 410, 421, 423, 425
cosubordination 58, 314, 330
countable 143, 144, 367
counterfactual *xiv*, 247, 389, 393, 395
culminatory telic *See telic*
Dardic 171, 453
“dative subject” 4, 148, 226, 327
ḍay / -ḍay 58, 84, 87, 99, 119, 124,
126, 128, 139, 164, 194, 254, 255,
333, 340, 381
decadal 190
default 251, 266, 294
deictic 12, 60, 64, 76, 77, 78, 88, 89,
90, 92, 106, 111, 116, 117, 143, 157,
171, 172, 185, 402
definite / indefinite *xi*, *xxi*, 1, 25, 79,
90, 96, 138, 142, 143, 144, 150, 151,
178, 179, 180, 181, 182, 245, 246,
247, 267, 276, 294, 344, 345, 367
ḍelki / ḍhelki / *D(h)elki* *Kharia* *See*
Kharia
demonstrative *xi*, *xxi*, 24, 44, 45, 52,
53, 61, 64, 66, 67, 73, 107, 112, 115,
135, 136, 137, 138, 142, 143, 171,
173, 174, 175, 182–187, 197, 408,
411, 414
dental 20, 29, 31, 33
departive *xii*, *xxi*, 127, 307–308
depictive *xiv*, 149, 207, 331, 389,
396–403, 434, 458
derived (morpheme) / derivation (with
respect to parts of speech) *x*, *xxi*, 1,
5, 17, 24, 42, 68, 69, 72, 73, 75, 76,
77, 79, 92–106, 107, 108, 109, 116,
126, 127, 128, 192, 202, 205, 214,
215, 216, 230, 262, 264, 268, 293,
298, 299, 300, 317, 333, 334, 371,
412, 453
derived voice 259
determiner 135, 137, 138, 148, 162,
183, 210, 414
Devanagari 19, 20, 21, 22, 23, 29
 devoiced 33, 240
Dhangar 5
Dhelki *Kharia* *See Kharia*
dialect(al) 4, 6, 71, 72, 76, 77, 96, 99,
103, 104, 105, 111, 112, 117, 122,
124, 148, 154, 195, 235, 253, 265,
266, 267, 296, 311, 312, 317, 324,
337, 449, 455

- differential (functions) xii, 259,
 260–276, 277, 279, 281, 282, 285,
 286
 diphthong ix, 21, 28, 32, 35
 direct case *See case*
 direct object *See object*
 direct participation 275, 276
 direct reflexive 288, 289, 291, 292,
 295, 298
 direct speech 405
 discourse 73, 318, 382, 386, 387, 455,
 456
 distal xxi, 156, 173, 183, 184
 distinctive 21, 27, 35, 36, 232
 distributive 102, 103, 192
 double causative *See causative*
 DP 75
 Dravidian 5, 6, 324, 454, 455
 dual(ity) xxi, 27, 65, 158, 159, 160,
 163, 169, 171, 173, 212, 247, 272,
 273, 355, 360, 378
 -*du?* 69, 248, 333, 334, 417, 418, 419,
 423
 Dudh Kharia *See Kharia*
 durative 69, 131, 215, 267, 278, 281,
 286
 dynamic xiii, 76, 374, 377, 378, 379,
 381, 388

 East Asia 206
 echo-vowel 20, 21, 22, 30, 39
 echo-word x, xxi, xxii, 25, 75, 118,
 129–133, 219, 333
 emotion middle 290
 enclitic *See clitic*
 endocentric 12, 75, 110
 English xvii, 5, 9, 10, 15, 18, 19, 22,
 51, 55, 56, 71, 74, 75, 76, 82, 85, 88,
 89, 92, 98, 99, 100, 102, 109, 112,
 118, 130, 136, 142, 188, 206, 254,
 308, 309, 341, 343, 353, 355, 357,
 358, 376, 381, 386, 388, 402, 406,
 428, 434, 435, 454, 457
 entity 57, 76, 85, 91, 93, 99, 101, 112,
 117, 139, 140, 141, 142, 144, 147,
 155, 158, 159, 160, 161, 162, 163,
 169, 173, 183, 185, 197, 209, 212,
 237, 238, 297, 305, 320, 352, 353,
 358, 359, 367, 374, 428, 434
 equational compound *See compound*
 equational predication 209
 equative xi, 202, 204–205
 ethnographic 9, 458
 ethnological 9

 Ethnologue 1, 6, 455, 456
 event 73, 77, 95, 96, 99, 101, 102,
 105, 206, 210, 215, 216, 217, 219,
 222, 230, 238, 240, 241, 243, 245,
 246, 247, 249, 250, 252, 253, 261,
 262, 266, 267, 268, 270, 271, 276,
 278, 284, 285, 288, 289, 290, 291,
 292, 295, 300, 301, 302, 303, 304,
 306, 310, 314, 316, 322, 324, 325,
 326, 330, 334, 345, 346, 358, 382,
 384, 390, 391, 394, 397, 414, 433,
 452, 456
 exception 6, 9, 14, 53, 61, 62, 65, 67,
 89, 94, 97, 98, 99, 100, 102, 104,
 105, 106, 110, 112, 116, 121, 137,
 143, 183, 191, 193, 194, 201, 209,
 214, 224, 234, 257, 278, 279, 280,
 287, 289, 322, 335, 357, 410, 414, 418
 excessive xii, xxi, 308–309
 exclusive xxi, 65, 160, 169, 212, 360,
 434
 expected (action) 161, 270–272, 276,
 452, 454
 experiencer xi, 4, 146, 225, 226–230,
 284, 327, 362, 364
 experiential perfect 205
 explicator verb *See v2*

 finiteness 10, 12, 314, 315, 316, 317,
 410, 424, 456, 457
 fully 210, 211, 239, 313, 314, 316,
 317, 390, 391, 395, 410, 411, 412,
 423, 424, 425
 non-finite xii, 25, 97, 150, 153, 239,
 313, 317–334, 396, 398, 424, 425,
 426, 454
 partial xii, 25, 107, 130, 200,
 313–317, 335, 411, 412, 423, 424,
 425
 fixed expression 127, 225–226, 314,
 438
 flap 29
 floating (clitics) 44, 45, 53, 54, 62, 66,
 67, 68, 293, 324
 focal 1, 68, 137, 194, 209, 257, 331,
 371, 391, 395, 398, 405, 426, 430,
 432, 433, 434
 focus xiv, xxi, 43, 53, 66, 72, 137,
 138, 142, 168, 179, 182, 311, 322,
 329, 331, 353, 370, 379, 399, 400,
 426, 427–438, 441, 452, 456
 formal 73, 96, 294, 348, 353, 415, 457
 free-standing 65, 94, 163, 168, 169
 fricative 29

- functional (morpheme / head) 12, 17, 36, 52, 58, 71, 77, 92, 109, 110, 112, 314, 322, 341
- future 4, 96, 105, 243, 245, 246, 247, 250, 253, 267, 276, 324, 334, 346, 380, 435
- gam*=*kon* 251, 328, 362, 404–405
- Gata? / Gta? 8, 92, 93
- geminate 31
- gender xi, 138, 139–142
- generic / genericity 215, 230, 264–268, 276, 278–279, 281, 282, 286, 299, 358
- genitive See *case*
- German xvii, 8, 9, 10, 376
- glide xxii, 28, 33, 34, 35, 143
- glottal (stop) 13, 20, 21, 22, 29, 30, 32, 34, 39, 93 256, 257, 340
- glottalization See *pre-glottalized*
- goal (semantic) 83, 144, 145, 158, 183, 398, 402, 430
- Gorum 8
- gradual telic See *telic*
- grammatical morpheme / marking 1, 12, 17, 25, 56, 59, 60, 61, 62, 69, 77, 78, 92, 117
- grammatical relations xiii, 25, 351–368, 454, 457
- Greek 376, 382
- grooming 269, 290, 291, 292
- grounded / grounding 77, 215, 377, 382
- Gta? See *Gata?*
- Gumla 1, 3, 6, 15, 16, 72, 76, 96, 112, 117, 265
- Gutob 7, 8, 93
- habitual 95, 96, 103, 104, 215, 239, 251, 264, 265, 266, 267, 268, 276, 333, 342, 380, 394
- =*har* (Sadri) 178
- hiatus xxii, 33, 35, 232
- hidden verb(s) 108, 111, 220
- Hill Kharia See *Kharia*
- Hindi 5, 6, 7, 10, 11, 13, 15, 19, 21, 23, 51, 99, 113, 116, 192, 193, 323, 324, 325, 333, 369, 401, 432, 439, 454, 456
- historical present 243, 244, 246, 247, 255, 378
- Ho 8, 103, 122, 454
- homorganic 21, 32
- honorific x, xi, xxi, 12, 17, 58–60, 71, 109, 160, 169, 209, 212–214, 215, 247, 355, 360, 377
- host x, 40, 52, 55, 61, 63, 66, 67
- human xxi, xxii, 120, 142, 158, 171, 173–178, 194, 195, 198, 199, 224, 336, 413, 450
- idiolect 103, 148, 253, 317, 337
- idiomatic 127, 429
- idiosyncratic 69, 101, 102, 126, 230, 231, 308
- imperative 247, 248, 249, 253, 257, 258, 316, 323, 334, 335, 338
- imperfective xi, xii, xxi, 105, 133, 207, 240, 241, 243–247, 259, 305, 330–333, 342–345, 393, 394, 397, 399, 400, 401
- imperfective converb See *converb*
- impolite See *polite*
- inalienable See *possession*
- inanimate See *animate*
- inceptive 254, 342
- inchoative 254
- inclusive xxi, 27, 65, 169, 170, 212, 360, 434
- incorporation x, 75, 118, 122–129
- increment(al) 256, 257, 305
- indefinite See *definite*
- indirect object See *object*
- indirect participation 275, 276
- indirect reflexive 289, 291, 292, 294, 298
- individuated 273, 274, 276, 278
- Indo-Aryan (IA) xxi, 4, 5, 7, 19, 21, 28, 29, 32, 33, 51, 98, 100, 113, 118, 122, 139, 150, 151, 190, 192, 193, 195, 201, 222, 224, 225, 235, 297, 298, 302, 311, 323, 326, 333, 334, 341, 342, 351, 377, 393, 394, 395, 408, 417, 423, 425, 440, 454, 457
- Indo-European 97, 318, 376, 382, 456, 457, 458
- Indological 23, 224
- inference xiii, 25, 349–350
- infinitive x, xii, xxi, 34, 60, 61, 66, 67, 85, 221, 312, 313, 321, 322–324, 333, 337, 341, 342, 346, 347, 348, 365, 366, 380, 389, 390, 391, 403, 415, 416, 417, 423
- infix See *affix*
- inflection(al) 17, 73, 262, 314
- information structure xiv, 25, 426–438
- inherent See *qualitative predication*
- instrument(al) xiii, xxi, 102, 111, 145, 146, 154, 320, 367, 389, 390, 410, 413, 414, 416, 417, 418, 419, 423

- intense / intensive xxi, xxii, 69, 87,
201, 206, 270, 308, 309
- intensity (phon.) 36, 46, 47
- intermediate agent / causer 235, 236,
367, 368, 413
- internally headed 407
- interrogative xi, xiii, xxii, 24, 25, 48,
77, 138, 178–179, 180, 181, 368–373,
408, 434
- interview xviii, xxiii, 11, 13, 15–16,
54, 72, 91, 112, 113, 116, 129, 183,
199, 231, 233, 234, 235, 265, 268,
283, 287, 296, 340, 355, 356, 357,
372, 441
- intervocalic 20, 29, 31, 124, 139, 302,
303
- intonation 47, 49, 293, 369, 420, 421
- intransitive xxii, 12, 212, 224, 259,
261, 262, 270, 271, 276, 277, 279,
280, 282, 283, 286, 287, 289, 295,
297, 305, 357, 358, 359, 413, 432
- IP 75
- IPA 23, 29
- irrealis xii, xxi, 35, 64, 240, 244, 245,
247–249, 252, 253, 256, 257, 258,
259, 272, 302, 314, 315, 316, 317,
324, 334, 335, 338, 342, 393, 394
- irregular 33, 101, 123, 144, 192, 221,
232, 257, 413
- iterative xii, xxi, 244, 264, 266, 267,
268, 298, 300, 301, 333, 342, 380,
399
- jaw-* 123, 124, 393
- jeʔ* 44, 58, 79, 110, 132, 169, 170,
173–178, 183, 185, 194, 216, 219,
253, 258, 312, 338, 345, 379, 383,
395, 405, 411, 422
- Jharkhand xv, xvii, xviii, xix, 1, 2, 6,
10, 11, 13, 15, 76, 83, 105, 342, 381,
417, 432, 439, 453, 457
- Juang 7, 8, 10, 69, 93, 297, 298, 334,
456, 457, 459
- juxtaposition x, 118, 120, 121, 136,
219, 420
- kaɾ* 5, 44, 54, 59, 61, 80, 84, 86, 88,
89, 95, 105, 107, 109, 123, 132, 140,
145, 150, 152, 157, 160, 161, 166,
169, 171, 172, 173–178, 183, 184,
191, 194, 202, 206, 216, 219, 226,
236, 248, 250, 252, 253, 272, 284,
285, 299, 300, 305, 310, 311, 321,
325, 327, 328, 340, 345, 356, 362,
363, 364, 366, 368, 370, 373, 374,
379, 384, 388, 395, 399, 400, 405,
408, 409, 411, 412, 415, 416, 418,
422, 427, 428, 431, 436, 441, 443,
444, 445, 447, 448, 450, 451
- Kharia (groups)
- d(h)elki* 5, 6, 9, 11, 458
- dudh, dūdh* 5, 6, 9, 11, 15, 19, 458
- Hill 5, 6, 9, 10, 453, 458
- Khasi-War 7
- Kherwarian 8
- Kolarian 6
- Korku 8
- Korwa 8
- Kurukh xviii, 5, 324
- labial 29
- length
- vowel 20, 21, 22, 31
- consonant 31
- lexeme 72, 135, 136, 210, 211, 335
- lexical x, 1, 4, 12, 18, 24, 42, 51, 52,
53, 56, 59, 68, 69, 72, 73, 74, 75,
77, 85, 87, 88, 90, 92, 97, 104, 106,
107, 109, 111, 112, 113, 117, 118,
124, 127, 129, 132, 137, 138, 139,
148, 153, 154, 157, 163, 173, 187,
196, 197, 219, 220, 226, 239, 257,
293, 297, 299, 302, 306, 307, 308,
309, 320, 321, 331, 341, 357, 393,
399, 401, 407, 408, 425, 455, 456,
457, 458
- lexicalization x, 68, 217, 261, 268,
458
- lexicon x, xvii, 9, 10, 12, 16, 24,
71–133, 139, 142, 456, 457, 458
- LH / low-high 24, 35, 36, 37, 38, 39,
40, 42, 44, 45, 51, 52, 54, 57, 60, 61,
68, 120, 178
- light verb 76, 77, 117, 224–225, 226,
227, 228
- linguistic area 4, 454, 456
- loan (word) 21, 28, 29, 31, 32, 77, 85,
98, 99, 100, 122, 131, 188, 201, 204,
224, 323, 341, 377, 391, 393, 441
- locative xxii, 77, 85, 88, 90, 92, 125,
143, 144, 145, 154, 156, 157, 158,
180, 183, 186, 206, 305, 320, 321,
367, 368, 410, 412, 413, 414, 416,
417, 418, 419, 423, 428, 436
- low-high See *LH*
- Maithili 377, 458
- Malaysia 7, 101

- manner (adverb) 72, 183, 216, 325, 326, 331, 401, 402, 403, 436
 masdar x, xii, xxii, 30, 42, 76, 92, 94–101, 104, 117, 121, 122, 123, 137, 149, 150, 200, 202, 215, 264, 265, 266, 267, 298, 299, 300, 301, 317–322, 327, 340, 397, 406, 412, 413, 414, 423, 424, 425, 426, 457
 mass (noun) 144, 367
 medial 66, 143, 173, 183, 184
 melodic overwriting 25
 methodology ix, x, 10, 13–16, 71–75, 351
 middle See *basic voice*
 minimal pair 27, 29, 30
 mobility See *clitic*
 modification xi, 25, 73, 138, 200–207, 323
 -*mon* 124, 227, 229
 Mon-Khmer 7, 101, 125, 456
 monomoraic 93
 monophthong ix, 21, 27–28
 monosyllabic 35, 36, 39, 40, 42, 44, 45, 62, 63, 87, 88, 90, 92, 93, 94, 97, 98, 99, 100, 102, 103, 122, 123, 126, 137, 200, 201, 202, 215, 231, 233, 234, 264, 265, 266, 300, 301, 318, 412
 mood 77, 209, 375, 377, 382, 383, 396
 mora(e) 92, 93
 morpheme-final 29, 33, 34, 64, 97, 240, 251, 256, 300
 morphology 13, 59, 79, 107, 351, 355, 371, 453
 morphophonology ix, 24, 33–35
 morphosyntax 13, 106
 motion 278, 279, 280, 285, 289, 290, 291, 292, 299, 302, 306, 389, 398, 401–403, 430, 456
 multiple semantic sub-bases xi, 215–219, 313, 330
 Munda xvii, 1, 4, 6, 7, 8, 9, 10, 11, 15, 75, 101, 106, 118, 124, 125, 178, 256, 259, 297, 322, 425, 426, 453, 454, 455, 456, 457, 458, 459
 North Munda 5, 8, 62, 103, 122, 143, 314
 Proto-Munda 5, 8, 92, 93, 314, 453
 South Munda xvii, 1, 7, 8, 9, 102, 103, 122, 124, 298, 322, 453, 457
 Mundari 5, 8, 62, 171, 344, 345, 359, 455, 457
 narrative (predication) 19, 209, 210, 375, 378, 379, 388
 nasal(ization) 21, 23, 27, 29, 32, 33, 101
 native xvii, 6, 28, 32, 33, 34, 51, 92, 94, 100, 103, 191, 192, 193, 195, 199, 233, 354, 378, 438
 natural reciprocal 288, 289, 290, 291, 292
 necessity xiii, 141, 348, 380, 415
 negate / negation / negative xiii, xxii, 25, 34, 58, 62, 67, 113, 149, 150, 178, 181–182, 211, 214, 249, 258, 312, 323, 329, 330, 334–341, 352, 373, 377, 378, 379, 400, 401, 424, 434
 Nepali 311, 312, 326, 349, 351, 456, 457
 Nez Perce 401
 Nicobar Islands 1, 7
 nominal xiii, 25, 72, 75, 89, 96, 118, 124, 318, 320, 351, 374, 382, 453
 nominal predication / sentence 19, 109, 209, 376, 385–388
 nominalizer 69, 75, 101, 104, 106, 221
 nominative 4
 nominative / accusative 212
 non-animate See *animate*
 non-endocentric See *endocentric*
 nonfinite See *finiteness*
 non-human See *human*
 non-individuated See *individuated*
 non-inherent See *qualitative predication*
 non-negated See *negate / negation / negative*
 non-overt See *overt*
 nonpast 245, 324
 non-periphrastic See *periphrastic*
 non-restrictive See *restrictive*
 non-sudden See *sudden*
 non-thorough See *thorough*
 non-volitional See *volitional*
 North Munda See *Munda*
 noun See *parts of speech*
 nominal phrase / noun phrase / NP x, xi, 1, 4, 12, 18, 25, 52, 66, 71, 73, 75, 106, 107, 108, 109, 110, 112, 117, 135–207, 212, 351, 359, 360, 365, 367, 382, 457
 nucleus 28, 33, 92
 number (gramm.) x, 12, 13, 17, 25, 37, 38, 52, 55–57, 58–60, 67, 71, 89, 109, 110, 112, 135, 138, 139, 140, 141, 142, 148, 158–162, 163, 187, 201, 209, 212–214, 338, 339, 353, 360, 367, 430

- numeral(s) xxi, 14, 102, 188–194, 195, 197, 199, 200
- nV*- x, 69, 92, 93, 94, 101–105, 192, 414
- object
- general / unspecified (includes direct objects) xiii, xiv, 1, 39, 40, 43, 52, 53, 66, 83, 94, 109, 124, 125, 128, 129, 142, 143, 144, 145, 154, 227, 238, 268, 273, 274, 276, 277, 278, 282, 294, 295, 304, 305, 317, 320, 321, 322, 358, 359, 367–368, 373, 398, 404–406, 420, 422, 423, 426, 427, 428, 429
 - “indirect object” 1, 142, 143, 144
 - physical 71, 72, 73, 75, 81, 82, 85, 99, 186
 - syntactic 60
- obligation xiii, 257, 309, 348–349, 380
- oblique See *case*
- onset 29, 32, 33, 34, 256
- ontological 73, 75, 77
- opaque 60, 103, 124
- open (class) 12, 77, 78, 90, 92, 97, 98, 117
- operator 314, 329, 330
- optative xii, xxii, 209, 210, 239, 240, 257–258, 334, 335, 338
- Oraon 5
- ordinal (numerals) 190–192, 193
- Orissa xv, 1, 2, 3, 5, 6, 16, 76, 99, 151, 152, 267
- Oriya 5, 6
- orthography / orthographical ix, 19–24
- outwardly directed 268, 269, 276
- overt 1, 72, 76, 79, 88, 93, 108, 111, 141, 142, 148, 294, 330, 331, 333, 336, 348, 352–359, 360, 361, 363, 364, 365, 367, 396, 397, 408, 428
- palatal 23, 29, 35
- partial(ly) finite(ness) See *finiteness*
- participle xii, xxii, 4, 97, 100, 219, 324, 330, 333–340, 361, 417–419
- parts of speech x, 10, 12, 15, 18, 24, 71–207, 317, 454, 455, 457
- adjective ix, xi, 1, 12, 18, 19, 72, 73, 90, 92, 99, 117, 136, 139, 200–207, 225, 396
 - noun ix, x, 1, 12, 18, 19, 71, 72, 73, 75, 89, 90, 92, 96, 97, 98, 99, 100, 102, 104, 106, 107, 108, 109, 110, 112, 115, 116, 117, 118, 124, 128, 136, 139, 144, 221, 225, 318, 357, 376, 377, 408, 453, 454, 457
- pronoun 12, 89, 107, 138, 139, 142, 168–182, 183, 408, 458
- verb ix, x, xi, xiii, 1, 4, 10, 12, 18, 19, 25, 52, 58, 59, 66, 71, 72, 73, 75, 76, 77, 79, 80, 90, 92, 96, 97, 98, 100, 104, 106, 107, 108, 110, 111, 112, 116, 117, 125, 128, 209, 212, 220, 221, 222, 224–230, 262, 293, 302, 318, 324, 335, 341–349, 351, 357, 375, 376, 377, 382, 388, 402, 453, 454, 455, 456, 457, 458
- passive xii, xxii, 6, 25, 45, 94, 221, 230, 259, 264, 288, 289, 291, 292, 293–298, 317, 318, 340, 367, 368, 413
- passive middle 289, 291, 292, 298
- past xi, xxii, 27, 33, 34, 59, 96, 117, 213, 215, 240–243, 244, 247, 252, 256, 259, 265, 266, 267, 268, 276, 281, 299, 300, 305, 314, 315, 334, 342, 344, 345, 380, 394
- Past II xi, xxii, 113, 209, 210, 239, 240–243
- patient 102, 212, 261, 267, 268, 294, 368
- perfect x, xii, xxii, 34, 35, 60, 61, 67, 210, 211, 213, 239, 240, 249–257, 293, 318, 342
- past 113, 239, 242, 252–253, 256
- present 209, 210, 239, 249–252, 256, 257
- irrealis 240, 252, 253–254, 257, 393
- pluperfect (Santali) 302
- periphrastic xiii, 25, 210, 211, 214, 241, 247, 337, 341–349
- person (grammatical, excluding references to the expression “TAM/PERSON-syntagma”) x, xi, 12, 17, 25, 33, 34, 35, 52, 58–60, 62, 63, 64, 65, 66, 67, 71, 77, 88, 93, 109, 110, 138, 160, 163, 169, 171–178, 209, 210, 211, 212–214, 215, 243, 245, 246, 247, 248, 249, 256, 257, 258, 272, 273, 293, 294, 299, 314, 316, 317, 322, 331, 335, 338, 339, 340, 348, 353, 355, 360, 367, 376, 377, 378, 388, 411, 424, 425, 457
- phoneme / phonemic 20, 21, 22, 27, 29, 30, 31, 32, 33, 143, 232
- phonetic 21, 28, 30, 129, 453
- phonology ix, xxiii, 8, 10, 24, 27–49, 52, 453

- phonotactic 13, 34, 62, 63, 248, 256
 phrasal affix x, 51–52, 57, 59, 60, 61
 phrase 12, 18, 52, 53, 57, 60, 61, 71,
 72, 108, 109, 148, 149, 216, 382,
 408, 420, 453
 physical object See *object*
 pitch 24, 35, 36, 37, 39, 42, 44, 45,
 46, 47, 48, 51, 52, 54, 57, 60, 61, 68,
 120, 178
 pivot 351, 359, 360, 362, 365, 367, 420
 place name 28
 plosive 20, 21, 29, 31, 32, 33, 34
 plural(ity) xxii, 18, 27, 55, 65, 87, 89,
 140, 141, 153, 155, 157, 158, 159,
 161–162, 163, 169, 170, 171, 173,
 193, 197, 201, 212, 213, 214, 247,
 272, 273, 294, 305, 335, 353, 360,
 378
 polar question xiii, 48, 49, 296, 368,
 369–370
 polite(ness) / impolite(ness) 160, 309
 Polynesian 401
 polysyllabic(ity) 30, 35, 87, 92, 94,
 96, 98, 99, 112, 137, 200, 215, 264,
 265, 267, 318, 331, 412
 portmanteau 209, 239, 258, 274, 275
 possession 147, 380
 alienable 148, 162
 inalienable x, xi, xxii, 25, 37, 52,
 57–58, 67, 100, 110, 112, 147,
 151, 162–168, 178, 293
 predicative 148, 380
 postalveolar 23, 29
 postfield 358, 426, 430
 postnominal xiv, 407, 408, 421–422,
 423
 postposition See *adposition*
 posture / postural 278, 280, 285, 289,
 290, 291, 292
 Praat 35
 pragmatic x, xiv, 1, 44, 45, 52, 53–54,
 66, 73, 126, 129, 163, 293, 318, 340,
 400, 414, 426, 427, 430, 433–438,
 455
 precategorical(ity) x, 24, 74, 75–92, 97,
 100, 106
 predication (for *narrative predication*
 and *qualitative predication*, see under
 the respective heading) 19, 73, 74,
 117, 209, 355, 382, 388, 396, 401
 primary 94, 100, 334
 secondary 94, 97, 317, 434
 predicative (function) 12, 17, 18, 19,
 71, 73, 74, 75, 76, 77, 78, 79, 80, 81,
 82, 84, 85, 86, 87, 89, 90, 91, 96, 97,
 100, 102, 103, 106, 108, 109, 112,
 117, 120, 124, 125, 130, 148, 175,
 179, 201, 202, 220, 225, 227, 265,
 266, 287, 317, 320, 334, 351, 371,
 372, 380
 predictability 74, 76, 84, 85, 90, 101,
 104, 123, 127, 232, 293, 308, 334,
 427, 436
 prefield 358, 426
 prefix See *affix*
 pre-glottalized ix, 20, 21, 22, 31–32,
 231, 232, 256, 257
 pre-nasal 27
 pronominal xiv, 333, 407, 408,
 410–419, 423, 424, 425, 426
 preposition See *adposition*
 present (tense / time) xi, xxii, 27, 34,
 209, 240, 243–245, 246, 247, 251,
 254, 255, 257, 259, 331, 334, 336,
 337, 342, 345, 348, 374, 375, 377,
 378, 379, 380, 388
 present perfect See *perfect*
 present progressive See *progressive*
 presentational 427, 432
 primary predication See *predication*
 proclitic See *clitic*
 proform xi, xxi, 11, 12, 25, 56, 76,
 77, 78, 88–90, 92, 106, 116, 117,
 138, 141, 147, 149, 150, 151, 163,
 168–178, 295, 352, 355, 363, 370,
 406, 411, 428, 431, 434
 progressive xii, xxii, 34, 240,
 245–247, 252, 255, 256, 257, 259, 342
 pronoun See *parts of speech*
 proper name 83, 147, 149, 150, 151
 property 69, 202, 209, 334, 364, 374
 propositional attribution xiv, 25, 333,
 389, 406–426
 prosodic / prosody ix, 4, 24, 35–49,
 52, 54, 60, 61, 396, 399, 458
 Proto-Austro-Asiatic 101
 Proto-Munda See *Munda*
 prototypical 10, 73, 166, 276, 277,
 278, 288, 289, 457
 proximal 183
 pseudo-compound See *compound*
 psyche (verb) xi, 226–230
 purpose xiii, 321, 389–390

 qualitative (predication) ix, xiii,
 xxii, 19, 25, 34, 76, 97, 209, 210,
 245, 256, 257, 336, 337, 340, 348,
 374–388

- inherent 209, 336, 374, 375, 377, 378, 379
 non-inherent 209, 256, 336, 374, 375, 377, 378, 380
 quale 91
 quantifier xi, 25, 52, 80–81, 107, 112, 138, 141, 174, 182, 187–194, 195, 196, 197, 198, 199, 202
 Raigarh 1, 3
 Ranchi xviii, 11, 13, 14, 15, 16, 83, 250, 328, 362, 453, 454, 456, 457, 458
 -*ray* See *day*
 recipient 144, 145, 364, 367, 368, 428, 429
 reciprocal xi, xxii, 17, 35, 42, 69, 128, 215, 230, 237–239, 274, 275, 288, 289, 290, 291, 292, 317, 318, 454, 456
 reduced x, 65, 148, 281
 reduplicate / reduplication x, xxii, 10, 17, 42, 75, 87, 88, 90, 92, 93, 94–101, 102, 104–105, 117, 121, 122, 123, 124, 137, 200, 201, 202, 215, 264, 265, 266, 300, 301, 318, 322, 399, 402, 412
 reference / referent(ial) 12, 17, 18, 19, 60, 71, 72, 73, 74, 75–92, 96, 97, 100, 102, 106, 109, 110, 115, 117, 118, 120, 128, 135, 140, 142, 160, 168, 169, 171, 173, 174, 175, 176, 177, 186, 191, 194, 195, 198, 199, 202, 210, 249, 317, 355, 358, 363, 364, 365, 431, 450, 455, 456
 relative clause xiv, xxii, 12, 25, 95, 110, 136, 175, 207, 319, 353, 365, 389, 392, 406–426, 454, 455
 reflexive xii, xxii, 25, 171, 221, 230, 237, 259, 268, 269, 288, 289, 291, 292, 293–298, 317, 318, 363, 364, 365, 367
 Remo 7, 8, 103, 122, 314, 315, 322, 455
 remote 96, 215, 266, 267, 268, 276, 281
 repetition xxii, 25, 87, 96, 131, 162, 201, 202, 296, 333, 399, 449
 restrictive xxi, 137, 406, 407, 421, 434
 resultative 218
 retroflex 20, 23, 29, 31, 33
 role (status) 83–84
 Roman (alphabet / script) 19, 22
 Romance 401
 root ix, 17, 24, 51, 52, 56, 59, 60, 92, 216, 314, 315, 325
 Sabar 6
 Sabbar 5, 6
 Sadami 4, 455, 457
 Sadri 4, 5, 6, 19, 21, 28, 33, 34, 51, 55, 69, 99, 100, 113, 116, 131, 178, 183, 186, 188, 189, 190, 192, 193, 195, 197, 199, 201, 204, 221, 223, 224, 225, 260, 295, 302, 324, 325, 326, 334, 349, 363, 377, 378, 381, 391, 395, 408, 417, 441, 442
 Sanskrit 21, 393
 Santali 8, 62, 93, 146, 302, 306, 307, 314, 457
 Sar 6
 Saur 6
 Savar 6, 453
 Savara 5, 6
 Sayar 6
 scalar 202, 204, 305, 336
 scene-setting 426, 428, 431, 433
 scope (prag.) 53, 57, 58, 60, 215, 218, 236, 329, 432
 secondary predication See *predication*
 self-directed 268, 269, 276, 281
 semantic base / head (use of term) 17; (form) 71–132, 135–143, 149–207, 214–239
 semantic map xii, 288, 298
 (semantic) shift 82, 90, 122
 Semelai 101, 456
 semel-iterative xii, 298, 300–301, 308
 Semitic 401
 sequential See *converb*
 shift See *semantic shift*
 Simdega xviii, 1, 3, 4, 6, 15, 16, 71, 74, 76, 77, 82, 92, 96, 108, 111, 112, 117, 136, 265, 266, 267, 287, 385, 417, 439
 simultaneous (events) 216, 326, 330, 397, 399, 400
 Sinhala / Sinhalese 171, 455
 size xi, 138, 139–142
 Sora 6, 8, 102, 298
 South Asia(n) 4, 23, 25, 100, 129, 153, 170, 171, 219, 226, 293, 324, 326, 327, 454, 456, 457, 459
 South Munda See *Munda*
 Southeast Asia(n) 125, 206, 453, 458
 Spanish 401
 speaker-specific 27, 144, 213, 231, 272, 274, 275, 287

- specific 1, 128, 270, 294
 spontaneous 33, 49, 230, 286, 290,
 291, 292
Sprachbund 4, 10
 standard (in comparison) 202, 203
 standard(ized) 5, 8, 9, 20, 21, 24, 29,
 242, 257, 324, 379, 439, 441, 449,
 451, 452
 stative xiii, 76, 97, 158, 348, 374, 375,
 377–381, 382, 384
 status (role) 83–84
 stem ix, xxii, 17, 24, 51, 52, 56, 59,
 60, 91, 129, 377
 stimulus 4, 226, 227, 228, 229, 230
 sub-base xi, 214, 215–219, 236, 272,
 313, 314, 398, 400
 sub-event 314, 316
 subject xi, xiii, xiv; (subject markers)
 212–214; (with “psyche-verbs”)
 226–230; (grammatical relation)
 351–367
 subordination xiii, 25, 389–406, 454
 successful (action) 260, 274, 275, 276,
 281, 284, 300
 sudden (event) xii, xxii, 270–272, 276,
 306–307
 suffix See *affix*
 Suir 6
 Sundargarh 1, 3
 superlative xi, 202–204
 suppletion / suppletive 262–264, 336,
 375, 377
 Surguja 1, 3
 Swiri 6
 syllable structure ix, 24, 32–33, 378
 syntactic atom x, 51–69, 109, 121,
 352, 356

 TAM/PERSON-syntagma (use of term) ix,
 18–19
 telic
 anticipatory xii, xxi, 62, 67,
 302–304
 culminatory xii, xxi, 31, 302–304,
 306
 gradual 131, 305
 other / general 66, 91, 110, 201,
 264, 293, 305, 310, 311, 312, 318,
 402, 403
 temporal xiv, xxii, 77, 87, 88, 90, 92,
 143, 145, 180, 183, 186, 206, 324,
 329, 334, 374, 377, 383, 389,
 390–393, 398, 410, 412, 413, 414,
 416, 417, 418, 419, 423, 428, 436

 thorough 269–270, 276
 thwarted (attempt) 274–275, 276, 281
 T'in 425, 455
 tone 35, 36, 44, 48
 Tongan 18, 454
 topic(al) xiv, 1, 142, 171, 172, 329,
 353, 357, 358, 360, 361, 362,
 426–438, 455, 456
 totality xii, xxii, 10, 182, 304–306,
 457
 transformative 254
 transitive xxii, 12, 201, 212, 224, 259,
 261, 262, 268, 270, 271, 272, 276,
 277, 279, 281, 282, 286, 287, 305,
 358, 359, 367, 368, 428
 translational active 289, 292
 translational motion 289, 290, 291,
 292
 transliteration ix, 19–24, 28, 29

 uncertain future 96, 267
 underived See *derived / derivation*
 unexpected See *expected*
 unmarked 95, 96, 135, 140, 144, 147,
 156, 158, 159, 171, 172, 173, 176,
 187, 189, 206, 215, 239, 242, 245,
 249, 264, 265, 266, 267, 268, 270,
 272, 273, 274, 276, 281, 286, 289,
 294, 299, 308, 348, 370, 383, 416,
 427

 v2 xii, xxii, 4, 10, 11, 25, 45, 62, 63,
 66, 68, 69, 80, 110, 116, 210, 211,
 257, 264, 293–313, 317, 318, 335,
 348, 356, 367 (see also under
 individual v2 entries)
 valency 230, 279, 317
 velar 23, 29
 verb See *parts of speech*
 verbalizer 79, 80, 106, 110, 116, 221
 Vietnam 7
 Vietnamese 206
 vigesimal 190
visarg 21
 voiced 20, 31, 33
 voiceless 63
 volitional / non-volitional / volitionality
 (See also *avolitional*) 224, 230, 270,
 271, 276, 277, 278, 279, 280, 281,
 282, 283, 284, 285, 286, 287, 309
 VP 73, 75

 Warlpiri 351, 458
 West Bengal 1, 3

- wh-question xiii, 296, 368, 370–373
word
 clitic 42
 grammatical (See also *syntactic atom*) 51–69, 118, 293
 phonological ix, x, xxii, 24, 35–46, 48, 51, 52, 53, 61, 65, 66, 68, 69, 88, 92, 93, 100, 102, 103, 121, 122, 123, 136, 178, 183, 194, 202, 209, 230, 257, 293, 317, 324, 349, 370, 436
 written ix, xxi, 23, 24
zero (copula, derivation, marking, etc.) xiii, 18, 88, 93, 107–108, 110, 111, 143, 259, 355, 376, 382–388, 403

INDEX OF AUTHORS

- Abbi, A. xix, 6, 7, 10, 100, 113, 164, 326, 453
- Aikhenvald, A. 42, 51, 52, 61, 453, 454
- Anderson, G.D.S. xvii, xix, 7, 10, 92, 93, 201, 256, 298, 453, 457
- Anderson, S.R. 52, 453
- Andrews, A. 212, 353, 453
- Archer, W.G. 11, 453
- Baḡ, J. 10, 453
- Banerjee, G.C. 9, 295, 297, 453
- Barnes, J. 457
- Bashir, E. 171, 453
- Benveniste, É. 376, 382, 383, 453
- Berlin, B. 201, 453
- Bhagat, S. 6, 10, 453
- Bhattacharya, S. 122, 453
- Biligiri, H.S. xxiii, 9, 11, 12, 13, 27, 33, 34, 102, 122, 127, 128, 136, 139, 142, 171, 175, 183, 199, 200, 246, 256, 257, 408, 411, 414, 453
- Bisang, W. xix, 206, 453
- Bloomfield, L. 209, 454
- Boersma, P. 35
- Bril, I. 259, 267, 454
- Broschart, J. 18, 454
- Catholic Biblical Association of America 220, 303, 342, 454
- Croft, W. 74, 454
- Delancey, S. 350, 454
- Deeney, J. 103, 122, 454
- Di Sciullo, A. 51, 60, 454
- Dixon, R.M.W. 51, 52, 212, 351, 453, 454
- Donegan, P. 4, 425, 454
- Druart, G. 455
- Durđun, J. 11, 439, 454
- Durđun, M. 10, 454
- Ebert, K.H. 4, 454
- Emeneau, M.B. 4, 454, 459
- Evans, N. 74, 455
- Falk, Y. 353, 455
- Fernandez, F. 103, 122, 314–315, 322, 455
- Filbek, D. 425, 455
- Floor, H. 9, 99, 201, 455
- Foley, W. 314, 351, 353, 455
- Gair, J. 171, 455
- Genetti, C. 324, 455
- Gheysens, V. 455
- Gonda, J. 259, 455
- Grierson, G.A. 5, 6, 455
- Grimes, B.F. 6, 455
- Haegeman, L. 262, 455
- Hengeveld, K. 72, 75, 357, 455
- Himmelmann, N.P. 396, 458
- Hoffmann, J. 344, 359, 455
- Hopper, P.J. 275, 278, 455
- Hovav See *Rappaport Hovav*
- Ishtiaq, M. 1, 455
- Jordan-Horstmann, M. 199, 221, 324, 455
- Kaufmann, I. 259, 455
- Kay, P. 201, 453
- Keenan, E.L. 351, 407, 455
- Kemmer, S. 259, 288–292, 298, 455
- Kerkettā, Kh.P. 11, 15, 60, 61, 78, 80, 88, 113, 114, 115, 116, 161, 162, 163, 178, 179, 190, 191, 192, 198, 199, 211, 216, 226, 244, 251, 254, 258, 295, 303, 307, 338, 342, 347, 353, 354, 355, 365, 391, 401, 403, 456
- Klaiman, M.H. 12, 259, 262, 278, 285, 456
- Koptjevskaja-Tamm, M. 396, 456
- Kruspe, N. 101, 456
- Kullū, A.V. 11, 15, 16, 439, 456
- Kullū, P. 5, 11, 220, 296, 303, 306, 342, 439, 440, 456
- Lambrecht, K. 427, 429, 433, 456
- LaPolla, R.J. 314, 330, 458
- Lazard, G. 350, 456
- Lehmann, C. 140, 456
- Levin, B. 262, 456
- Lewis, M. P. 1, 6, 456
- Lichtenberk, F. 237, 238, 456

- Maas, U. xix, 10, 36, 94, 97, 209, 318, 396, 456, 457
- Mahapatra, B.P. 10, 163, 456
- Malhotra, V. 5, 7, 9, 11, 12, 13, 58, 76, 113, 118, 119, 120, 121, 123, 130, 138, 139, 148, 163, 164, 171, 183, 192, 195, 199, 224, 235, 236, 326, 335, 336, 340, 373, 379, 408, 420, 456
- Masica, C.P. 4, 456, 457
- Matson, D.M. 7, 69, 334, 456
- Michailovsky, B. 350, 456
- Montaut, A. 113, 456
- Moral, D. 4, 456
- Narasimhan, Bh. 401, 456
- Neukom, L. 4, 146, 302, 307, 314, 454, 457
- Nichols, J. 262, 263, 457
- Nowrangi, P.S. 221, 349, 457
- Osada, T. xix, 74, 171, 455, 457
- Palmer, F.R. 351, 457
- Patnaik, M. 7, 457
- Peterson, D.A. 457
- Peterson, J.M. xvii, 10, 22, 36, 77, 94, 98, 101, 219, 221, 245, 302, 306, 311, 313, 324, 326, 342, 349, 350, 351, 375, 433, 457
- Pfeffer, G. 9, 439, 457
- Pinnow, H.-J. (including HJPa & HJPb) xxiii, 6, 7, 8, 9, 11, 16, 22, 27, 56, 80, 100, 101, 102, 103, 104, 113, 114, 118, 122, 125, 128, 130, 131, 140, 149, 158, 163, 167, 188, 189, 190, 191, 192, 194, 195, 196, 198, 199, 206, 217, 225, 227, 236, 242, 243, 251, 252, 253, 272, 295, 297, 298, 304, 309, 310, 311, 312, 314, 315, 316, 317, 326, 328, 329, 339, 340, 362, 370, 395, 400, 401, 411, 412, 415, 416, 417, 418, 440, 457
- Pustejovsky, J. 75, 91, 458
- Rappaport Hovav, M. 262, 456
- Rehberg, K. 10, 28, 32, 33, 36, 44, 46, 48, 120, 369, 458
- Rijkhoff, J. 72, 75, 357, 455, 457
- Risley, H.H. 6, 458
- Roy, S.Ch. / Roy, R.Ch. xxiii, 5, 9, 130, 315, 439, 458
- Sadock, J. 51, 458
- Sāhu, R.P. 11, 458
- Sasse, H.-J. 73, 351, 458
- Schultze-Berndt, E. 396, 458
- Simpson, J. 351, 458
- Sinha, D. 9, 458
- Stampe, D. xix, 4, 10, 425, 454, 459
- Talmy, L. 401, 458
- Tea Districts Labour Association 9, 455, 458
- Thompson, S.A. 275, 278, 455
- Upadhyay, V.S. 9, 458
- Van Valin, R.D., Jr. 314, 330, 351, 353, 358, 455, 458
- Vidyardhi, L.P. 9, 458
- Weerink, D. 35
- Williams, E. 51, 60, 454
- Yadav, R. 377, 454, 457, 458
- Zide, N.H. 7, 8, 10, 92, 93, 101, 201, 453, 457, 458