UC Santa Barbara
Himalayan Linguistics

Title

Language contact in Jharkhand: Linguistic convergence between Munda and Indo-Aryan in eastern-central India

Permalink

Journal

Author

Peterson, John

Publication Date

2010

DOI

Copyright Information

Peer reviewed

Himalayan Linguistics

Language contact in Jharkhand
Linguistic convergence between Munda and Indo-Aryan in eastern-central India

John Peterson

Universität Leipzig

ABSTRACT

The present study takes a closer look at language convergence in Jharkhand in eastern-central India, concentrating on Indo-Aryan and Munda languages. Although it is well-known that the Indo-Aryan languages which function as *linguae francae* in the region – such as Sadri, Bengali and Oriya – have had an enormous impact on the morphosyntax and lexicon of the Munda languages, in this study I call attention to a number of convergences which to my knowledge have so far gone unnoticed, many of which appear to originate in Munda, while others are of uncertain origin. These include, among others, the emergence of inalienable possession as a morphological category and incipient dual marking in the pronominal paradigm in Sadri, similarities in categories denoting 'from' and 'to' or 'begin' and 'keep on', as well as a number of interesting areal developments of the genitive, including 3rd person marking, focus marking, or becoming part of the copular stem in several languages of the region.

KEYWORDS

Jharkhand, Indo-Aryan, Munda, convergence, inalienable possession, genitive, ambiguous categories

Language contact in Jharkhand: Linguistic convergence between Munda and Indo-Aryan in eastern-central India[*]

1 Introduction

It is widely recognized that the Indo-Aryan (IA) languages have had an enormous impact on the Munda languages of eastern-central India, especially with respect to syntax and the lexicon. It is also generally assumed that the Munda languages for their part have – at least in some distant past – affected the Indo-Aryan languages, although it has so far been extremely difficult to demonstrate this with any certainty. The present, exploratory study takes a different approach and attempts to demonstrate that convergence between Munda and IA continues to this day in Jharkhand in eastern-central India and that this influence operates in both directions, with the Munda languages often influencing IA in regions where speakers of Munda languages are found in large numbers. The study concentrates on the development of the genitive in the various languages of the region and also on further phenomena which can only be explained by language contact.

The study is structured as follows: Section 2 presents a general introduction to the linguistic situation of present-day Jharkhand and previous work on convergence in this region. Section 3 presents a number of examples of linguistic convergence in Jharkhand, including the spread of inalienable possession as a grammatical category from Munda to IA (3.1), a single category expressing both inception as well as durativity/general imperfectivity (3.2), similarities between expressions denoting 'from' and 'to' (3.3), and an anticipatory telic category which is often closely linked to a morpheme denoting 'take' (3.4). The section ends with an example of change in

progress, the emergence of a dual category in Sadri (IA) under Munda influence (3.5). Section 4 is a discussion of the genitive in the IA and Munda languages of Jharkhand. We begin with the forms of the genitive (4.1) and then proceed to its further development in languages of both families to a nominalizer / (part of) the infinitive marker (4.2), the marker of the 3rd person singular (4.3), its use with copular forms (4.4), and finally its use as a focus marker (4.5). Section 5 presents a summary and outlook.

2 The languages of Jharkhand

The state of Jharkhand is one of the linguistically most diverse regions of the subcontinent and is home to languages from three families – Indo-Aryan (Indo-European), Munda (Austro-Asiatic) and Dravidian. The present study focuses on the first two of these families, i.e., Indo-Aryan and Munda: Due to the large number of languages spoken in the area (ca. 30) and the fact that many of the phenomena discussed in this study have often gone unnoticed in previous studies, I concentrate here on the IA and Munda languages of the region with which I am most familiar: Santali, Mundari and Ho, the three major North Munda languages, Kharia, the only South Munda language spoken in Jharkhand, and Sadri, as the representative of IA, due to its special status as the traditional *lingua franca* in central and western Jharkhand.[1] This study draws upon published works and complements these with data which I have collected during six research trips to Jharkhand.

Map 1 shows the position of Jharkhand in South Asia. The state to the north of Jharkhand is Bihar, of which Jharkhand used to form the southern half; the state to the east of Jharkhand is West Bengal, that to the west is Chhattisgarh, while the state to the south of Jharkhand is Orissa. From a historical and cultural perspective, the regions in these last three states which border the present-day state of Jharkhand may be considered to be part of "larger Jharkhand", and a considerable number of Munda languages are also spoken there.

---

[1] Although Dravidian languages such as Kurukh and Malto will eventually also have to be taken into consideration, they will not figure prominently in the present discussion for a number of reasons, including above all the availability of data for the relevant features. Where I do have such information, generally from Masato Kobayashi (p.c.), this information will be included. However, as Dravidian seems to be the "newcomer" to the region and is geographically marginal to the regions I am most concerned with here, this gap, unfortunate though it is, should not affect our discussion significantly.

Map 1: Jharkhand and its neighbors²

The following presents a brief overview of the five major languages in this study.

2.1   *South Munda: Kharia*

Kharia is the only South Munda language spoken in Jharkhand, the only South Munda language for which Sadri is the general *lingua franca*, and the only South Munda language spoken in the direct vicinity of the North Munda languages, most notably Mundari. According to the Ethnologue (Lewis 2009), it was spoken in 1997 by ca. 292,000 people in India and by 293,580 in all countries. It is spoken primarily in southwestern Jharkhand and in the neighboring districts in the states of Chhattisgarh and Orissa.

Although as a South Munda language Kharia is clearly related to the North Munda languages, the divide between North and South Munda is considerable, both in terms of the lexicon as well as their morphosyntax, and the two branches are by no means mutually intelligible. Hence, North and South Munda will consistently be dealt with separately here.

2.2   *North Munda: Mundari, Santali and Ho*

Mundari, Santali and Ho are the three North Munda languages with the largest numbers of speakers. As with North Munda languages in general, these three are linguistically so closely related that they are largely mutually intelligible. The Ethnologue gives the following figures with respect to the number of speakers (Lewis 2009): For Santali and Mundari (in 1997): 5,960,000 (Santali) and 1,550,000 (Mundari) in India and 6,170,900 (Santali) and 1,560,280 (Mundari) for all countries. For Ho (in 2003): 1,500,000 speakers. Generally speaking, the Mundari-speaking heartland is to the east of the Kharia-speaking area, although the two languages are both spoken in many of the same villages, while the main Ho-speaking area is somewhat further to the east. The

---

² The original map, which has been modified here somewhat, is copyrighted material by sa PlaneMad/Wikipedia, and is freely available from the following website:

http://de.wikipedia.org/w/index.php?title=Datei:India_Jharkhand_locator_map.svg&filetimestamp=20081229064837

main Santali-speaking areas are in the eastern parts of the state of Jharkhand and adjoining areas of the states of West Bengal and Orissa.

2.3   Indo-Aryan: Sadri

Sadri is the IA *lingua franca* for much of western and central Jharkhand, and is also used on a daily basis by speakers of a number of Munda and Dravidian languages. It has an array of alternate names, of which "Sadani" is the term generally used outside of India to refer to this language. I will refer to this language here as "Sadri".[3] According to the Ethnologue, in 1997 there were 1,965,000 speakers of Sadri in India, with 2,165,000 speakers in all countries (Lewis 2009).

Sadri varies considerably in terms of morphosyntax depending on the native language of the person speaking it, and the texts I have from three different ethnic groups (Sadri *jait*) differ in a number of respects, one of which will be presented in section 3.5. There are also considerable differences in pronunciation; although these are clear signs of ethnic identity for the speakers involved, they will not be dealt with further here as the present study concentrates on morphosyntactic features.

2.4   Multilingualism in Jharkhand

The level of bi- and multilingualism is relatively high in Jharkhand, at least in southwestern Jharkhand. In my own experience, in this region most speakers of Munda languages are fluent in their native language as well as Sadri and Hindi. Speakers of Sadri, on the other hand, are generally fluent in Hindi as well but seldom have active knowledge of a Munda language, although this is not entirely unknown. Finally, speakers of Santali, which is spoken further to the east, are often more familiar with Bengali than with Hindi or Sadri, and Bengali has had an enormous impact on Santali, comparable to that of Sadri on the Munda languages of western Jharkhand. Thus, although individual levels of multilingualism certainly vary greatly, there is a relatively high level of multilingualism in Jharkhand, so that convergences between the many languages of this region are to be expected.

With respect to morphosyntax, Kharia, Mundari, Ho, Santali and Sadri all share the following general characteristics:

➢ All five are predominantly predicate-final ("SOV"), although not rigidly so;
➢ All five show a strong head-final tendency in general, not only with respect to the position of the predicate but also internal to clause-level constituents ("NPs") as well as with respect to the position of bound morphemes;
➢ All five have predominantly agglutinating / enclitic grammatical marking;
➢ All five share the following traits:

---

[3] The Ethnologue lists the following alternative names: Chota Nagpuri, Dikku Kaji, Ganwari, Gauuari, Gawari, Goari, Jharkhandhi, Nagpuri, Nagpuria, Sadan, Sadana, Sadani, Sadari, Sadati, Sadhan, Sadhari, Sadna, Sadrik, Santri, Siddri, Sradri (Lewis 2009).

Although this language is generally referred to in western studies as *Sadani*, I will not follow this practice here, as *Sadani* in the region itself refers to a number of closely related languages, such as Panch Parganiya, Kurmali, Khortha and Sadri, all of which are spoken by the "Sadan", i.e., the Indo-Aryan-speaking groups of western and central Jharkhand. Hence, the term *Sadri* will be used here throughout.

- <u>Finite predicates ("verbs")</u>[4] are marked for TAM and the person/number/honorific status of the subject (S/A). In Santali, Mundari and Ho, the predicate also marks for the object (P), but not in Kharia or Sadri.
- <u>Referential expressions ("NPs")</u> – none of the five languages has grammatical gender, and contentive morphemes used in reference ("nouns") all have one invariable form, to which enclitic case and number markers attach.

Many of these structural characteristics apply to most languages of the subcontinent while others, such as a single, invariant form for "nouns" and the lack of grammatical gender, are perhaps more typical of eastern South Asia. Although this in itself perhaps points to a prolonged period of close contact between the two language families, the present study will concentrate on more tangible points of convergence between IA and Munda in Jharkhand which are not found elsewhere in South Asia or only rarely.

A note is also in order here on terms such as *Sprachbund* or "linguistic area": When these terms are encountered in this study, they merely serve as abbreviations for clumsier expressions such as "a region in which a number of languages are found which can be shown to have undergone some degree of convergence, regardless of the direction of this convergence." In other words, I will not provide any definition of "linguistic area" in terms of a minimum number of shared traits limited to this region or some similar definition but will only refer to individual patterns of convergence which I have noticed. In addition to the exploratory nature of the present study, this is also due to the fact that, in my opinion, "linguistic areas" in any meaningful sense of this term do not exist, i.e., languages do not form linguistic unions (*Sprachbünde*) but merely affect one another more or less gradually in situations of prolonged contact, such as in Jharkhand.

Furthermore, attempts to count similarities often overlook areas in which the grammars diverge, often quite unexpectedly if one assumes that a *Sprachbund* exists. For example, as will be mentioned in the following section, almost all languages of Jharkhand have numeral classifiers, which would normally be taken as a sign of convergence (a view which I accept). However, this tells us little about the behavior of these markers, e.g., that these classifiers, in addition to their facultative use with numerals in all languages of the region, may also directly follow a noun in Sadri to denote definiteness/specificity, as in eastern IA in general, while this is not possible, e.g., in Kharia. Hence, the question necessarily arises as to where the cut-off point for a *Sprachbund* should be drawn, i.e., how similar must two languages be in order to form a linguistic area? Which traits are most important? And do differences count as negative similarities, or are they merely ignored? Hence, the present study will not assume the existence of "linguistic areas" and terms such as this, when they are used, are for the sake of brevity only.

---

[4] The topic of parts of speech in Munda languages is quite complex, and it is debatable whether lexical classes such as "noun", "verb", "adjective", etc., are applicable to many of these languages. As this issue is beyond the scope of the present study, I will simply refer here to predicates and referential expressions in Munda languages without further comment. Where I feel that this may be unclear, I will include by way of explanation "V", "NP", etc., in quotation marks. This has no further significance and is only meant as an aid to the reader. On the other hand, parts of speech are clearly distinguishable in Sadri so that I will speak here of nouns, verbs and adjectives.

2.5   Previous studies

The present study is not the first to point out that the languages of Jharkhand have converged over the course of time. Two earlier studies which deserve special mention are Abbi (1997), which concentrates on IA influence on Dravidian (Kurukh) and Munda (Kharia), and Osada (1991), which deals with general convergence tendencies among all three families. The following list, although not exhaustive, summarizes the major convergences noted by these two authors:[5]

- **Lexicon**. As Abbi (1997) notes, there has been massive borrowing from IA into both the Munda and Dravidian languages of Jharkhand. This borrowing goes far beyond the mere designation of new concepts and objects for which the borrowing language had no designation and includes, among others, numerals, kinship terms, body parts, natural objects, temporal organization, adjectives, and verbs.
- **Sequential Converbs ("conjunctive participles")**. As Abbi (1997) points out, the presence of sequential converbs in the Munda and Dravidian languages of Jharkhand in all likelihood derives from IA. Consider, e.g., Kharia, where two of the three converbal markers, =*ke* and =*ker*, are borrowed directly from Sadri and the third, =*kon*, derives from the Kharia form *ikon* 'make, do', apparently a calque of the Sadri form, as at least the converbal marker =*ker* in Sadri derives from *kʌr*- 'make, do' (Peterson 2006: 243f.).
- Other aspects of the Munda and Dravidian languages of Jharkhand noted in Abbi (1997) which have been influenced by IA include the **phonological inventories** of these languages, which have accommodated non-native phonemes into their inventories, various **grammatical morphemes**, the **correlative clause construction** typical of so much of South Asia and which was already well developed in Old Indo-Aryan (OIA), as well as **numeral classifiers**, a trait also noted by Osada (1991).

In addition to numeral classifiers, Osada (1991) notes the following traits shared by all three language families of the region:

- **Echo-word formation**, generally termed "melodic overwriting" in general linguistics, involving forms such as Mundari *jhaka-maka* 'shining with gold, silver or tinsel', Sadri *jhaka-maka* 'shine, shining', Kurukh *jhakā makā*[6] 'shining', Kharia *jhaka maka* 'shining' (Osada 1991: 101). This trait is common throughout much of South Asia.
- **Onomatopoetics**. Although this is common throughout much of South Asia, it has been noted elsewhere (e.g., Tiwari 1960: xliii) that at least some of these forms in IA perhaps derive from Dravidian and Munda languages.

---

[5] At this point I would like to stress once again that the purpose of this study is not to demonstrate that Jharkhand is a *Sprachbund* with clearly defined external boundaries but rather an area with a number of linguistic convergences. Hence, although many of the following points are true not only of Jharkhand but also of much of South Asia (and beyond), I nevertheless list these as possible points of convergence of the languages of Jharkhand to one another (and possibly to languages outside of Jharkhand). To what extent Jharkhand itself may be linguistically defined by these and other traits will be discussed briefly in Section 5.

[6] Printed as *jhaka maka*. Unfortunately, Osada (1991) abounds in typographical errors and the author has kindly provided me with a hand-corrected version of this text, which I gratefully acknowledge here.

Finally, Osada (1991) calls attention to a trait which is common to apparently all of the languages of Jharkhand and which will occupy a prominent position in the discussion in the following pages:

➢ The four-way distinction in the present tense – and only there – between generally suppletive **existential** and **identity copulas** on the one hand and **negative** and **non-negative copular paradigms** on the other. Although this would not seem to be especially interesting from a cross-linguistic perspective, and although the existential/identity distinction (or some other, similar distinction) is common to many languages of the subcontinent, the presence of this distinction together with negative and non-negative paradigms, generally resulting in four suppletive stems in the present tense, is sufficiently seldom in South Asia to warrant mention. These forms will be dealt with in more detail in section 4.4.

3 Some examples of linguistic convergence in Jharkhand

A number of convergences among the languages of Jharkhand can be added to the features given in Section 2 as indicative of intensive and prolonged language contact between Munda and IA in Jharkhand. To begin with, there are a number of features which, while perhaps interesting from a South Asian perspective, are typical of general cross-linguistic tendencies. For example, in a number of languages of Jharkhand we find that the word for 'want' derives from a word meaning 'seek, search for', e.g. Kharia *lam* 'search for, seek; want', Sadri *khoj* and Kurukh (Dravidian) *bedd-* (Grignard 1924: 142; Masato Kobayashi p.c., 2009), with the same meanings. Furthermore, in many languages of the region, an expression meaning 'place' has developed to become a general allative/locative marker in structures closely resembling compounds, cf. Kharia *maṛa boʔ=te* [cave place=OBL] 'at the cave' and Sadri *bās ṭhin* [bamboo.plant place] 'at the bamboo plant'.

Such semantic developments may appear to be rather unexpected if one compares them with Hindi, however the fact that these are highly common developments in languages from around the world detracts considerably from their usefulness in demonstrating convergence in Jharkhand. As such, the following concentrates on developments which are not only uncommon in much of South Asia but are also either cross-linguistically uncommon or, as in the following section, whose source appears to be clear.

3.1 Inalienable possession

Virtually all Munda languages distinguish morphosyntactically between alienable and inalienable possession. The following presents examples of inalienable and alienable possession in Kharia (1) and Santali (2).

Kharia (Peterson 2006: 123):

Inalienable	Alienable	
(1) *aba=ɲ*	*iɲ=aʔ*	*khoṛi*
father=1SG	1SG=GEN	village
'my father'	'my village'	

Santali (adapted from Neukom 2001: 32)[7]

Inalienable	Alienable	
(2) *hɔpɔn=me*	*am=ak'*	*oɽak'*
son=2SG	2SG=GEN	house
'your son'	'your house'	

In the IA languages of central India, on the other hand, it is highly unusual to find a morphosyntactic distinction between alienable and inalienable possession. Nevertheless this distinction is found in Sadri: Alienable possession is marked by a genitive attribute which precedes the head noun (possessum), while inalienable possession is marked by the enclitic form *=har*, although only for a 3rd person possessor, to which we return below.

Consider the following two examples. (3) also provides an example of both alienable and inalienable possession: *bhai=har=mʌn=ʌk* modifies *sewa* as a genitive attribute in alienable possession. On the other hand, *bhai=har=mʌn=ʌk* itself is marked for inalienable possession, as *=har* denotes that *bhai* is inalienably possessed by a sister or brother:

Sadri

(3) *bhai=har=mʌn=ʌk*   *sewa*
    brother=3POSS=PL=GEN   service
    'his/her brothers' service'

(4) *bhʌuji=har=mʌn*
    sister.in.law=3POSS=PL
    'his/her sisters-in-law'

Grierson (1903: 280) refers to this marker as a marker of "definiteness" and connects it to the homophonous marker in neighboring Chhattisgarhi (IA). However, speakers I consulted, both L1 and L2 speakers of Sadri, confirmed that this marker is used to denote a possessive relationship which, in both my own data as well as in the texts in Jordan-Horstmann (1969), is restricted to body parts, kinship terms and a few other terms denoting people, such as *sʌŋgi* 'friend'.

Since inalienable possession is found so seldom as a grammatical category in central IA but is found in virtually all Munda languages, it seems reasonable to assume that this trait has been "borrowed" from Munda into Sadri. This is supported by the form and function of *=har* itself: *=har* may only refer to a 3rd person possessor – there is no special construction to denote inalienable possession with a 1st or 2nd person possessor, and possession of either semantic type is expressed by means of a genitive attribute with these possessors. This functional distribution receives a natural explanation when one considers the origin of this marker; *=har* originally denoted definiteness and was later reinterpreted to denote inalienable possession, a closely related concept (cf., e.g., Taylor 1996).

---

[7] Glosses from other sources have been adopted to the conventions used in this study. Morphemes which I consider to be enclitics (in all languages) will be indicated here by the sign "=" instead of "-", which is reserved for affixes. This is not always in line with the description of these elements by other authors, however these modifications do not directly affect our discussion and only serve to unify the presentation of the data.

Definiteness, or perhaps more appropriately, specificity (cf. Neukom and Patnaik 2003: 24ff.), is marked in many eastern IA languages among other means by attaching a classifier, which otherwise primarily follows numerals, directly to the noun itself. This is also true of Sadri. Consider the following example, where *lʌɽki=go* refers to a girl who has already been mentioned in the preceding story:

Sadri

(5) | *u=kʌr* | *bad* | *dhan=mʌn=ke* | **lʌɽki=go** | *bʌrka-l-ʌk* | *sijh-ʌl-ʌk* | *sukha-l-ʌk …* |
|---|---|---|---|---|---|---|
| that=GEN | after | paddy=PL=OBL | girl=CLASS | boil-PST-3SG | boil-PST-3SG | dry(TR)-PST-3SG |

'After that, **the girl** boiled the rice paddy, dried [it] …'   [BCB.0071][8]

Thus, *=har*, which presumably derives from such a classifier, would have yielded a definite / specific interpretation when attached directly to a noun in Sadri. This explains why *=har* only indicates a 3rd person possessor: 1st and 2nd person possessors were presumably marked by a genitive attribute (which simultaneously marked the NP as definite) when *=har* began to be reanalyzed, as they are today. Thus, as *=har* developed further from a marker of specificity into a marker of inalienable possession, this was restricted to 3rd-person possessors.

In sum, although the exact details of this development await further study, it is clear that the CATEGORY of inalienable possession in Sadri has been "imported" from Munda, but making use of morphology which was already found in Sadri.[9]

3.2 'begin' and 'keep on/IPFV'

One and the same morpheme can sometimes have both an inceptive interpretation ('begin') and either a durative interpretation ('keep on') or that of general imperfectivity in languages of both Munda and IA stock in Jharkhand and further to the east into Bengal.[10] Cf., e.g., the morpheme *laʔ* in Kharia, which derives from Sadri *lag-* 'begin': Although *laʔ* in Kharia most commonly denotes general imperfectivity (6) (despite its etymology), it can also denote the inception of an action or event (7), especially one of prolonged duration.

---

[8] Examples given with sources of this type refer to data from my own texts.

[9] In an interesting aside: Although the morphosyntactic category of inalienable possession in Sadri derives from Munda, Sadri also exerted influence on at least one Munda language with respect to this marker: Cf. Sadri *ke* 'who?', which has the alternative form *ke=har* [who=3POSS] 'who?', apparently with no difference in meaning. Similarly Kharia *ber* 'who?' with its alternative form *behar* < **ber=har* 'who?', again with no apparent difference in meaning. Although the original functions of these alternative forms are unclear, the two extended interrogative forms are clearly related to one another, and the fact that Kharia makes use of the Sadri marker *=har* suggests that this has been borrowed there from Sadri. Hence, the category of inalienable possession, which was motivated by Munda, has now found its way from Sadri back into Munda, but in a very different function.

[10] Although the data are still somewhat preliminary, the ambiguity dealt with in this section does not seem to depend on the *Aktionsart* of the contentive morpheme ("verb"), at least in those languages for which I have been able to check this, although this may have originally played a role in the emergence of this ambiguity. Further research is necessary.

Kharia

(6) … *khaɽiya   lebu=ki   pujapaʈh   <u>karay=na   laʔ=ki=may</u>, …*   [AK, 2:6]
       Kharia   man=PL   sacrifice   do=INF   IPFV=MID.PST=3PL
       '… the Kharia men <u>used to perform</u> sacrifices … '

(7) *hobne=te=ga   ubar   koleʔ   kunɖuʔ jal=te   bajhe=kon   "ʈ̃ay ʈ̃ay"   <u>toroʔɖ=na</u>*
       that.much=OBL=FOC   two   parrot child   net=OBL   get.trapped=SEQ   "tay, tay!"   cry=INF

   <u>*laʔ=ki=kiyar*</u>.   [BB, 2:33]
   "IPFV"=MID.PST=3DU
   'Meanwhile (= in that much), two baby parrots got caught in the net and <u>began crying</u> "Tay!
   Tay!".'

   Mundari has a rather similar category; here the morpheme *–jan* is used with certain intransitive predicates such as *sen* 'go', *nir* 'run away', *apir* 'fly away' and *hoka* 'cease, stop' to denote that the subject "went, ran or flew away and is not yet returned, or that an action has been stopped for the moment and has not yet been resumed, though the intention of resuming it exist[s]." (Hoffmann 1905[2001]: 183) With transitive and other intransitive predicates this morpheme denotes 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." (ibid, emphasis in original). In other words, *–jan* denotes both the inception of an event or action and also its (perhaps interrupted) continuation. Thus, despite the very different marking we find rather similar categories in both North and South Munda.[11]

   A similar category is also found in at least some Indo-Aryan languages of the region, e.g., in Sadri:

Sadri

(8) *rait bhe–l–ʌk.   ʌb buɖha buɖhiya   bicar   <u>kʌr-ek hel-l-ʌẽ</u>   ki...*
       night become-PST-3SG now old.man old.woman   thought do-INF "begin"-PST-3PL CMPL
       'It became night. Now the old man and the old woman <u>began thinking</u> …'
                                                                              (Jordan-Horstmann 1969: 129)

---

[11] Although this does not appear to be discussed in any grammar of Santali, my own research suggests that Santali once also had this category. Altogether I have come across four morphemes, *sɛn* 'go', *hɛc'* 'come', *bɔlɔ* 'enter' and *calaɔ* 'go', which behave irregularly in that they form the past tense not with the "past middle marker" *-en*, as do most other intransitive predicates, but rather with the zero-marked "non-past active", which is otherwise generally only found with transitive predicates and otherwise always denotes a non-past action or event. With these morphemes, the "past middle marker" *-en*, which is cognate with Mundari *-jan*, denotes that the speaker is unsure of the action/event being narrated, and with 1st persons it can also imply that the speaker was drunk, as s/he cannot recall whether the action/event in fact occured.

   If we assume that this category in Santali once expressed the notion more generally of an action which was begun and is still in progress (or perhaps more generally: not completed), then I believe that this explains not only the semantics of these four morphemes in Santali – especially their use with 1st person subjects – but also why this class so closely resembles the small class of intransitives noted by Hoffmann and referred to in the text above. Further research is necessary.

Although it is clear from the context (not given here) that *hel-* in (8), which as a full verb has the meaning 'swim (trans. [sic!]); to be busy, occupied' (Jordan-Horstmann 1969: 175), is to be interpreted as 'begin', all speakers I consulted agreed that, without the first sentence in (8) and without ʌb 'now', (8) could also have the meaning 'The old man and the old woman <u>kept on thinking</u>.' Similarly, further to the east, the auxiliary verb *lag-* in Bengali, when followed by an infinitive, can mean both 'begin' and 'keep on' (Radice 1994: 240, note 16). This "dual" use (at least from an English-speaking perspective) of a single category seems to be restricted to Jharkhand and Bengal, although much more work is necessary on this topic.

3.3   Similarities between 'from' and 'to'

In at least some Indo-Aryan languages of the region, the same marker can be used with an ablative sense as well as with an allative sense. This meaning will be referred to here as "extensional", as it refers to the temporal or physical distance extending to or from an action / event or location. I will gloss it as 'EXT'.

<u>Sadri</u>

(9) *se=khʌn   le   hʌmre=mʌn=ke   cik   bʌɽaik kʌh–ʌl ja-t=he.   aiȷ   le.*
    that=time   EXT   1NSG=PL=OBL   Chik   Baraik   say-PTCP PASS-IPFV=PRS.3SG   today   EXT
    '<u>Since</u> that time we are called "Chik Baraik". <u>Until</u> today.'

Similarly, cf. the Bengali postposition/noun <abadhi> [ɔbodhi] 'since; until; n. limit'.

This feature may ultimately derive from Old Indo-Aryan: For example, in Sanskrit in the dual and plural (but not in the singular) the dative and ablative cases are regularly expressed by the same form. Furthermore, there is the preposition *ā* with the meaning 'up to', which takes a noun in the ablative, cf. *ā samudrāt* [up.to ocean.ABL.SG] 'up to the ocean' (Coulson 1976: 94); although the meaning of *ā* is unambiguous, the fact that it requires the ablative is noteworthy.

The Munda languages of the region clearly distinguish the ablative and allative meanings, although they do show a morphosemantic connection between the two. Cf. Kharia *tay* 'from', which is also found in the allative postposition *khoʔtay* 'up to' (< **khoʔ tay* [place ABL] 'from the place'). Similarly, Santali: *həbic'* 'up to; until', *həbic'khɔn* 'from; since', where *khɔn* is otherwise the ablative postposition; the situation in Mundari is similar: *həbiʔ* 'up to; until', *həbiʔkhɔn* 'from; since' (Ganesh Murmu p.c., 2009).

3.4   The anticipatory telic marker

A further possible candidate for IA/Munda convergence concerns an *Aktionsart* category which I have termed the "anticipatory telic" in earlier studies on Kharia. This category in Kharia denotes that one event comes to a conclusion (hence "telic") and that another event obligatorily follows it (hence "anticipatory").

Kharia

(10) *tay   raja   jhaṛi… ḍokloʔ   remaʔ   ḍoṭh=oʔ   ro   ho=ki=te   "masih=te   ate*
     then king   all   meeting call   ANTIC=ACT.PST and   that=PL=OBL   Messiah=OBL   where

   jorme=na   ayiʔ j?"   gam=oʔ   ro   juŋ=oʔ.
   be.born=INF   PRS.COP   say=ACT.PST   and   ask=ACT.PST
   'Then the king <u>called</u> all … to a meeting <u>and</u> asked (= said and asked) them "Where is the
   Messiah to be born?"'   [Kullū 1992:2 = Luke, 2:4]

This category has a counterpart in the Santali "pluperfect" and "irrealis" marker *-le* (and variants), which has largely similar semantics. Consider (11).

Santali (adapted from Neukom 2001: 80f.)

(11) *jɛmɔn=e   bɔlɔ   gɔt'-len=a,   tɛmɔn=ge=kin   kiləp   ɛsɛt*
     as=3SG.SUBJ   enter   TEL-ANTIC:[MID][12]=IND   then=FOC=3DU.SUBJ   close   close

   gɔt'-ked=e=a.
   TEL-ACT.PST=3SG.OBJ=IND
   'The moment he <u>got in</u>, they closed and shut him up.'

Similar functions are also found for *-le* in Mundari and Ho (cf. Osada 2008: 126; Deeney 1975: 39, respectively).
   Sadri has a predicative category which is similar in many ways to the anticipatory categories just described in North and South Munda and which is also marked by *-le*.

Sadri

(12) *…jolha   u=ke   poṭra-le   ʌur ghoṛa upʌr bʌiṭh-le   rʌ<i>h   ge-l-ʌk.* [13]
     Muslim   that=OBL   grab-ANTIC and horse   above sit.down-ANTIC remain<LNK> go-PST-3SG
   '… the Muslim <u>grabbed it and</u> remained seated (= <u>sat down [and]</u> remained) on top of the
   horse.'   (Nowrangi 1956: 165)

(13) *…ek   tawa   ḍabk-a-e   de-le   bʌdʌk-i   uṭh-en.*   [SBT.0023]
     one   griddle boil-CAUS-LNK   give-ANTIC hurry-LNK   stand.up-PRS.3PL
   '… <u>if you boil</u> them on a griddle, they will scatter here and there (= hurry).'

A similar form/function combination is found further to the east in Bengali, with the "conditional participle" marked by the suffix *-le*, and also to the south in Oriya, with similar forms (cf. Neukom and Patnaik 2003: 251f.). Although Neukom and Patnaik refer to this form as the "conditional converb", they also explicitly note that it is found in other functions as well, such as

---

[12] Neukom inadvertantly glosses *-len* here as active.

[13] *ge-* 'go' in (12), *de-* 'give' and *uṭh-* 'stand.up' in (13) are so-called "vector verbs" which denote *Aktionsart* and follow the lexical stem marked by the linker *-e*. Their exact function is not of direct relevance here.

habitual situations where one action/event/situation precedes another (cf. their examples (89), (90) and (95)).

What is especially interesting with respect to this category is that the marker in Kharia, *ɖoˀɖ*, is homophonous with the contentive morpheme *ɖoˀɖ* 'take', from which it derives. This is relevant since the root 'take' in Sadri is *le-*, which might suggest that the Santali, Mundari, Ho and Kharia forms are borrowings/calques from IA.

From a linguistic point of view, the anticipatory marker *-le* in Sadri, as also in Bengali (Chatterji 1926: 1004, §736), does not derive from a verb meaning 'take' but rather from the past participle marker *-l* followed by the locative marker *-e*. Furthermore, it is not restricted to anticipatory events but is also found in a number of other environments, most notably conditionals, as in (13), although there are of course (often) semantic similarities to an anticipatory interpretation. Nevertheless, its similarity to the Santali, Mundari and Ho forms as well as the semantics of the etymological source of the Kharia construction suggest at least the possibility that the form which is historically analyzable as *-l-e* [-PTCP-LOC] has been reinterpreted as the Sadri verbal stem *le-* 'take' and borrowed in this function into Kharia and North Munda.

However, there is also strong evidence against assuming that the form *-le* has been borrowed into North Munda from IA: An *-l* is found in many Munda languages to denote the "anterior" or "cislocative" (cf., e.g., Pinnow 1966: 141, §3.2.12.1.2; Anderson 2007, §§4.1-4.2) and there are also further differences, both semantic as well as the fact that the IA forms are obligatorily nonfinite, whereas the Munda forms can form part of the marking on a finite predicate. Further research is necessary. At any rate, the presence of these markers, all with similar functions (and forms) at the very least suggests convergence throughout much of this region, regardless of the etymology of the individual markers.[14]

3.5   Under (re-)construction – change in progress

This section calls attention to what seems to be a case of incipient contact-induced change, namely the appearance of a dual category in the Sadri pronominal system, which is being introduced by bilingual Munda speakers, in this case speakers of Kharia.

Table 1 provides an example of the Sadri pronominal system as determined in conjunction with a native speaker (L1) of Sadri. Table 2, by way of contrast, presents this system as determined in conversations with a Kharia L1 speaker, i.e., L2 speaker of Sadri (ca. 50 years old), although she has been an active speaker of Sadri since a very early age (all forms appear in the direct or "nominative" case).

A number of differences between the two systems immediately become apparent. To begin with, the Kharia-L1 system has only two levels of politeness and only in the 2nd persons: unmarked and honorific. The Sadri L1-system, on the other hand, has four levels of honorificity in the second persons as well as two in the 3rd persons.

But what is even more apparent is the fact that the Kharia L1-system has an OBLIGATORY dual category which is not present in the L1-system. This demonstrates how new categories can be gradually transferred from one language to another. The dual forms can be analyzed as follows: *dui* means 'two', whereas *=jhan* and *=o*, like *=go* in Table 1, are classifiers, *=jhʌn* being restricted to

---

[14] There are two morphemes, *xacc-* 'rip' and *liŋgh-* (meaning as an independent morpheme unknown) which convey a similar meaning in Kurukh (Dravidian) (Masato Kobayashi p.c., 2009; Grignard 1924: 147). In view of the comments given above, a connection to Munda and IA languages seems likely. Further research is necessary.

humans, whereas *=o* and *=go* are compatible with all nouns. The Kharia L1-speaker I worked with insisted that the dual forms in Table 2 are obligatory in Sadri and that the use of the plural forms for two individuals is incorrect. It is of course well known that speakers often believe they speak differently than they in fact do, nevertheless her comments make it clear that in her opinion this is the correct usage and is to be observed in "good" Sadri, i.e., at least in careful speech.

Person	Singular	Plural
1	*mõe*	*hʌmre*
2		
Familiar unmarked	*tõe*	*tohre*
Familiar polite	*tohre*	*tohre=mʌn*
Polite	*rʌure*	*rʌure=mʌn*
Honorific	*ʌpne*	*ʌpne=mʌn*
3		
Unmarked		
Proximal	*i(=go)*	*i=mʌn*
Distal	*u(=go)*	*u=mʌn*
Honorific		
Proximal	*i(=mʌn)*	*i=mʌn*
Distal	*u(=mʌn)*	*u=mʌn*

Table 1: The Sadri pronominal system (Sadri L1 speaker; note: *=go* 'CLASS', *=mʌn* 'PL', *dui* 'two')

Person	Number		
	Singular	Dual	Plural
1	*mõe*	*hʌmre dui=jhan* *hʌmre duy=o*	*hʌmre(=mʌn)*
2			
Unmarked	*tõe*	*tohre dui=jhan* *tohre duy=o*	*tohre(=mʌn)*
Honorific	*raure*	*raure dui=jhan* *raure duy=o*	*raure(=mʌn)*
3			
Proximal	*i*	*i dui=jhan* *i duy=o*	*i=mʌn*
Distal	*u*	*u dui=jhan* *u duy=o*	*u=mʌn*

Table 2: The Sadri pronominal system (Kharia L1 speaker; note: *=o* 'CLASS', *=jhʌn* 'HUM.CLASS', *dui* 'two')

L1-Sadri speakers, on the other hand, ACCEPTED the dual forms with no hesitations, as they do not violate any principles of Sadri and simply mean 'we both', 'you both' and 'they both', respectively. However, they responded that this category is not obligatory and that the use of the plural for two individuals is perfectly acceptable.

The differences between the two systems become understandable when one considers the Kharia "pronominal" system, presented in Table 3. Although the Kharia system differs from the Kharia-L1 system in Sadri in a number of ways, the categorial distinctions between the Sadri-L1 and Kharia-L1 systems of Sadri receive a natural explanation when one takes the data in Table 3 into consideration: Kharia has only two distinct honorific levels, in all persons – honorific and unmarked.[15] This distinction is retained in the Kharia-L1 system of Sadri, although it is lost in the first and third persons. Furthermore, Kharia has an obligatory dual category, or rather, a dual which is obligatory in the first and second persons (number marking is never obligatory in the 3rd person). Hence, this category has been carried over by this speaker into Sadri by entirely grammatical, Sadri-internal means.

If non-Kharia-L1 speakers notice anything at all, then probably only that the Kharia tend to be very exact in denoting the number of individuals; that at least some of these speakers consider this category obligatory in Sadri is likely to go unnoticed. Given enough time, such a system, which explicitly denotes whether one, two or more individuals are involved in a specific situation, could eventually become obligatory, as transparency is generally assumed to be preferred in multilingual situations.

		Singular	Dual / Honorific	Plural
1 INCL		*iɲ*	*anaŋ*	*aniŋ*
1 EXCL			*iɲjar*	*ele*
2		***am***	*am=bar*	*am=ɟe*
3	Human	Non-human		
PROXIMAL	*u=kaɽ*	*u=jeʔ*	*u=kiyar*	*u=ki*
MEDIAL	*ho=kaɽ*	*ho=jeʔ*	*ho=kiyar*	*ho=ki*
DISTAL	*hin=kaɽ*	*hin=jeʔ*	*hin=kiyar*	*hin=ki*

Table 3: The Kharia "pronominal" system (simplified)

4 The Genitive

We now turn to the genitive, discussing both its forms and the functions which it has assumed in the languages of the area. These developments provide us with a unique window on the dynamics of language contact in Jharkhand and beyond. We will now deal with these developments individually.

4.1 The genitive marker in the languages of Jharkhand – Form and distribution

Depending on its environment, the genitive in Sadri has the following forms: *=kʌr / =ʌk / =k* (adapted from Jordan-Horstmann 1969: 45). At present, it is not clear to what extent, if any, there are semantic and/or pragmatic differences between *=kʌr* and *=(ʌ)k* where both forms are found in "free" alternation.

---

[15] In the first person, the (otherwise) exclusive form *iɲjar* is used to denote honorificity.

– After nouns ending in –*a* → =*kʌr* or =*k*:

(14) *ghoḍa=k   /   ghoḍa=kʌr*
    horse=GEN   horse=GEN
    'the horse's; of the horse'

– After nouns ending in a consonant → =*kʌr* or =*ʌk*:

(15) *bap=ʌk   /   bap=kʌr*
    father=GEN   father=GEN
    'the father's; of the father'

– After =*mʌn* 'PL', =*har* '3POSS' → =*ʌk*:

(16) *bhai=har=ʌk   sewa*   *bhai=mʌn=ʌk   sewa*
    brother=3POSS=GEN   service   brother=PL=GEN   service
    'his/her brother's service'   'the brothers' service'

– Elsewhere (rare): =*kʌr*:   *mirtu=kʌr* 'death's; of death'

       The form =*(ʌ)k*, found in Sadri, is also found in a number of Munda languages (accommodated to Munda phonotactic rules):[16]

(17) Santali =*ak'* [aʔ]   Mundari =*aʔ*   Kharia: =*aʔ*

       There are also a number of alternate genitive forms in North Munda, cf. Santali –*reak'* [reaʔ], used with an inanimate possessum; this form is analyzable etymologically as the locative marker –*re* and the "original" genitive marker =*ak'* [aʔ]. Other forms include –*ren* with an animate possessum, and occasionally –*reaŋ*, which is found in the titles of stories (Neukom 2001: 29f.). The form with pronouns is =*ak'*, cf. *ac'=ak'* [he=GEN:INAN] 'his' (adapted from Neukom 2001: 36). Similar (but slightly different) comments hold for Mundari and Ho.
       In Kharia, the genitive has the basic form =*aʔ*, with the allomorphs =*ʔ*, =*yaʔ* and =*waʔ*, largely depending on the form of the stem-final segment of the preceding unit. As this alternation is basically phonotactically determined and –*y*– and –*w*– primarily serve to avoid a hiatus, I will speak in the following merely of the form =*aʔ* for Kharia as well as for Santali, Mundari and Ho.
       /aʔ/ is thus likely the "original" form of the genitive in all four Munda languages mentioned here (and other North Munda languages of Jharkhand), and there are three possible explanations for the similarity of this form to =*(ʌ)k* in Sadri and other IA languages of the region: 1. The form =*aʔ* derives from Proto-Munda and =*(ʌ)k* in Sadri has been borrowed from Munda; 2. The form =*(ʌ)k* from Sadri (and a few other eastern IA languages such as Maithili) was borrowed into the Munda languages of Jharkhand, and 3. Sadri =*(ʌ)k* and Munda =*aʔ* are both native elements in their respective families and the similarity between them is purely coincidental.

---

[16] With non-borrowed morphemes in Kharia an underlying /k/ or /g/ in the coda is realized as [ʔ] (Peterson 2008: 439). A similar situation holds in Mundari (Osada 2008: 101) and Santali (Ghosh 2008: 26; 30). Thus, we can safely assume an "underlying" form =*ak* (or =*ag*) for the genitive in Kharia and North Munda.

Although we will not discuss this issue here in any length, there is convincing evidence that these markers can indeed be accounted for in both language families without assuming borrowing from one family into the other, hence for the present study we will merely assume that this similarity, striking though it is, is purely coincidental. It will be dealt with in more detail in a later study. But whatever its origin, the genitive marker seems to have developed further in both Munda and IA in rather unpredictable but similar ways, to which we now turn.

4.2 From genitive to (part of the) infinitive marker / nominalizer

Let us begin our discussion here with Sadri. The first further development of the genitive in Sadri was its use to mark the infinitive. In this respect, Sadri differs from all dialects of Bhojpuri, of which it is often considered a dialect. In Bhojpuri, spoken to the northwest of Sadri, the cognate form is *-e* or *-ai*, apparently depending on the preceding segment.[17] The form in Sadri, on the other hand, is *-ek*. This form consists etymologically of two different markers, *-e* and *-k*:

➢ *-e* derives from the now semi-productive locative marker *-e* (Jordan-Horstmann 1969: 80), which is still found in a number of more-or-less fixed expressions in Sadri such as *ghʌr-e* 'at home'. Cf. also, further to the east, the Bengali infinitive in *-te*, which is homophonous with one of the locative markers *-e/-te*, from which it derives (cf. Chatterji 1926: §747).
➢ On the other hand, Jordan-Horstmann (1969: 79) writes that *-k* is "historically identical with the OIA affix *-ka* which is either pleonastic or serves to express diminutiveness", although in the same paragraph she herself calls this analysis into question (cf. also Chatterji 1926: §§430, 721). I also find this analysis unlikely as this diminutive form, although highly productive in earlier stages of IA, is otherwise entirely unknown in Sadri. Rather, in my opinion, this *-k* derives from a nominalizing function of the genitive, presumably via speakers of North Munda languages. We will now deal with this topic in some detail.

That there is a close relationship in many languages between the genitive and nominalization is well known. For example, in many languages of Nepal we find the (etymological) genitive serving as a nominalizer in a variety of functions (cf., e.g., Noonan 1997). There are also signs of such a connection between nominalization and the genitive in North Munda. For example, in Santali the genitive case marker can productively be added to contentive morphemes which denote an action or state to derive "patient nouns". Consider (18).

Santali (adapted from Neukom 2001: 58)
(18) *gɛt'* 'cut'   *gɛt'=ak'* 'the thing [which was] cut'
      poṇḍ 'be white'   *poṇḍ=ak'* 'the white thing'

---

[17] Unfortunately, Tiwari (1960) does not seem to deal with this marker in his grammar. Instead, the form he refers to as the infinitive, in *-l* (cf. e.g., Tiwari 1960: 187), appears to be a verbal noun, cf. *kaïl / karal*, which he refers to as the old and recent forms of the infinitive, respectively. On the other hand, the text specimens given at the end of his grammar contain (unglossed) forms such as *bigāre cahalī* '(they) tried to spoil' (p. 225), *na kare pãĩ* 'I cannot do' (p. 227), etc., which strongly suggests that the infinitive, in the sense that I use this term, is marked by *-e*. Shukla (1981), dealing with the northern-most Bhojpuri dialect, cites only the cognate form *-e*, also without any trace of the *-k*-element discussed in this section.

=*ak'* here is a "nominalizer" and behaves similarly to =*ic'* (cf. (19)), which seems to have no other function in the language than that of a "nominalizer" in reference to an animate entity (from the same source):

(19) *gɛt'=ic'*   'the person [who was] cut'

Similarly, Osada (1992: 63) refers to the nominalizing function of the "possessive" (i.e., the genitive) =*aʔ* in Mundari.

This "nominalizing" function is generally restricted not only to the 3rd person, as is to be expected, but also to the singular. Consider in this respect (20), in which the "nominalizer" is used to derive the so-called "independent possessive pronouns" from the unmarked possessive pronouns in Mundari,[18] but only in the singular.

<u>Mundari</u> (adapted from Osada 1992: 66)

(20) *aɲ=ag=aʔ*	*am=ag=aʔ*	*ay=ag=aʔ*
1SG=GEN=NMLZR	2SG=GEN=NMLZR	3SG=GEN=NMLZR
'mine'	'yours'	'his/hers'

Cross-linguistically, the genitive/nominalizer is often used to mark an infinitive. What makes the development in Sadri so interesting is the combination of the locative and genitive, i.e., -*e=k* [-LOC=GEN] > INFINITIVE, especially the relative position of the two markers: While it is quite common cross-linguistically for the genitive to form a kind of "oblique stem" to which other case markers and postpositions may attach, the reverse order, found in Sadri, is at best highly uncommon, and I am aware of this ordering in only one other language group – North Munda. Consider once again from the discussion in 4.1 the alternative form of the genitive in Santali with an inanimate possessum, -*re=ak'*, and also its Mundari equivalent -*re=aʔ*, both of which may be analyzed historically as '-LOC=GEN'.

With respect to the relative chronology of the two markers which combined to form the present-day Sadri infinitive marker, i.e., -*e* and -*k*, it is clear that -*k* is the younger of the two. This is shown by the fact that both Sadri and all Bhojpuri dialects, which are perhaps Sadri's closest linguistic relatives, have the form -<u>*ai*</u> or -*e*, while only the Sadri form contains the -*k*. Hence, -*k* is an innovation of Sadri, i.e., that "dialect" which is in close contact with North Munda. Furthermore, the fact that the South Munda language Kharia regularly incorporates Sadri verbs into its lexicon and consistently marks these with -*e* but never with -*ek* (cf. Peterson 2006, §6.3.3) also suggests that this borrowing pattern is the continuation of an earlier pattern in which Sadri verbs were borrowed in their infinitival form, a common source of borrowed verbs (cf., e.g., Wohlgemuth 2009: 80f.).[19]

---

[18] As such, I disagree with Osada's interpretation of this form as a nominalizer, as the forms it "nominalizes" are already assumed to be pronouns. For a more detailed discussion, cf. Peterson (2005: 402, fn. 13). However, this does not directly affect our argumentation here.

[19] This could also be a case of what Wohlgemuth (2009) refers to as the "reduced infinitive", in which the -*k* is dropped in all borrowings. However, considering the arguments given above, I argue that the apparent "reduction" in the case of Kharia is misleading and that the stem-forming suffix -*e* in Kharia is simply the continuation of an earlier -*k*-less infinitive.

Hence, the typological rarity of this order and its presence in North Munda (but not in Kharia) and Sadri suggests that this is a good example of linguistic convergence between IA and North Munda. With respect to the direction, the fact that no Bhojpuri dialect has this element suggests that this development originated in North Munda and spread to Sadri, although it could of course be a kind of "interlanguage phenomenon".

4.3 *From genitive to 3rd person, singular*

In a considerable number of languages in the region, the marker for the 3rd person in some predicative categories is homophonous with the marking of the genitive, namely *=ʌk, =ak* or *=k* in Indo-Aryan languages and *=(a)ʔ* in North Munda. In this study, we will assume that the genitive has evolved further in these cases, probably via a nominalizing function, to also assume this function.[20] However, even if this analysis is rejected, it is nevertheless noteworthy that in so many languages of the region the genitive is homophonous with the marker of the 3rd person, singular, at least in some categories, and this areal pattern must be accounted for.

In Sadri the genitive has come to mark the 3rd person, singular of the past tense. Table 4 presents the conjugation of *kha-* 'eat' in the past tense by way of example.

Person	Singular	Plural
1	*kha-l-ō*	*kha-l-i*
2	*kha-l-e*	*kha-l-a*
3	*kha-l-ʌk*	*kha-l-ʌẽ*

Table 4: Sample paradigm of the past tense in Sadri: *kha-* 'eat' (Nowrangi 1956: 93)

Sadri shares this trait with many eastern IA languages of Jharkhand such as Kurmali (Grierson 1903: 149), Sadri Kol (Grierson 1903: 159), Panch Pargarniya (cf. the texts in Grierson 1903: 168ff.) but also further to the north, such as Magahi (Verma 1991) and Maithili (Jha 1958; Yadav 2003).[21]

---

[20] For reasons of space, no systematic attempt will be made here to justify this decision, although eventually the details of such a development will have to be accounted for. For now, we will simply refer to the large number of languages from around the world – including Indo-Aryan – which have incorporated an erstwhile participle into their finite paradigms and also note the natural affinity between the 3rd persons, especially in the singular, and nominalizations.

[21] The form is also found in the 3rd person singular and plural of the subjunctive in Sadri, which functions as a kind of 3rd person imperative. Hence, the comments given in the following text also refer to these two forms. Interestingly, Jha (1958: 479f., §685) writes of the markers of the 3rd person, non-honorific containing a /k/ in Maithili that they are "of unknown origin", being possibly connected with Kashmiri *-h-* and *-k-* but also noting that they may be linked with the "indefinite pronoun" *keo*. I believe that all these forms – in all of the languages under discussion – can easily be accounted for if we assume that this marking derives from (a nominalizing function of) the genitive.

Chatterji (1926: 989f., §721) also notes that verb forms with *-k* are found in Bengali in the 2nd person past and future and non-honorific 3rd person past and future (and in some dialects also in the 3rd person, past, habitual), as well as in the 2nd person, imperative (rarely). He writes that this form there is considered archaic but also that "In the N[ew]B[engali] Standard Colloquial, however, <<-ka>> may be used with negatives in all persons and tenses as a detached word, without any special force, unless it be of some sort of finality … in [these cases] the genitive phrase may be taken as one group-word to which the affix is added." (Chatterji 1926: 990).

Similarly, in some North Munda languages, *=(a)ʔ* can be used to mark the 3rd person, singular. To begin with, the 3rd person, singular in Mundari is marked on the predicate by the enclitic forms *=e / =i* (depending on vowel harmony), which are the "original" markers for this person. However, the forms *=eʔ / =iʔ* are also found in certain environments (cf. Osada 1992: 64f.), which at least implies a similar structure as in the above-mentioned IA languages, i.e., forms marked by the "genitive". This is summarized in Table 5, based on the discussion in Osada (2008: 120).

	SG	DU	PL
1INCL	*=ɲ*	*=laŋ*	*=bu*
1EXCL		*=liŋ*	*=le*
2	*=m*	*=ben*	*=pe*
3	*=e/=i/=eʔ/=iʔ*	*=kin*	*=ko*

Table 5: Mundari enclitic person/number markers

Unfortunately, it is not entirely clear from the discussion in Osada (1992, 2008) just what principles, if any, guide the speaker/writer in choosing the form with or without the glottal stop. However, having examined all sentential examples in the main text of Osada (2008), I believe that the general tendency can be illustrated by example (21).

Mundari (Osada 2008: 106)
(21) *biŋ   coke=ʔ   jom-ja-ʔ=i=a.*
      snake   frog=3SG.SUBJ   eat-INGR-TR=3SG.OBJ=IND
      'The snake is eating the frog.'

Note that the enclitic subject marker, which precedes the predicate here,[22] is marked by the form with a glottal stop whereas the object marker, which is predicate-internal, is the form without the glottal stop. Although there were also exceptions, this explains the OVERWHELMING MAJORITY of forms encountered in Osada (2008).

In the closely related Ho, this connection between the 3rd person, singular animate subject and the erstwhile genitive has been completely grammaticalized, so that the form *=eʔ*, with the glottal stop, always marks animate 3rd person subjects while *-ī*, without the glottal stop, always marks animate 3rd person objects (Deeney 1975: 1 and especially 20). Santali also makes use of the genitive/nominalizer to express the 3rd person but to a much less extent; in addition to the copula

---

More research is necessary here to determine the exact distribution of these forms. This last comment, on the appearance of this form with negatives in colloquial Bengali, would seem to fit in well with the focusing function of this morpheme, cf. section 4.5.

[22] As a general tendency (although not a hard-and-fast rule), subject markers attach to the predicate when this is the only element of the clause but otherwise to the word which directly precedes the predicate if there is one, regardless of its function.

(see Section 4.4), it is only used to mark a 3rd person, singular, inanimate object in the imperfective and non-past applicative (Neukom 2001: 122).[23]

In sum, there is a clear connection between the 3rd person, singular and the "genitive / nominalizer" in Mundari and Ho, and to a lesser degree in Santali. This contrasts with the IA languages of the region, in which the erstwhile genitive/nominalizer *=ʌk* marks the 3rd person, singular on verbs, but only in the past tense (and a few other forms, see footnote 21).

4.4 From genitive to stem-internal element of the copula

We also find a *=k / =ʔ* in various parts of the paradigm of the copula in many languages of the region, which is again homophonous with the genitive, from which I suggest it ultimately derives. We begin with a discussion of these forms in the North Munda languages, starting with Santali. However, before proceeding to the copula, we must first discuss the principles of person marking in Santali (and North Munda) in general. Diagram 1 presents a (simplified) overview of the structure of the predicate in Santali (adapted from Neukom 2001: 61). A similar structure is also found in Mundari, Ho and other North Munda languages.

STEM + Conjugation + TAM + Voice + **OBJECT** + indicative marking + **SUBJECT**
            (Neutral,   (Active,
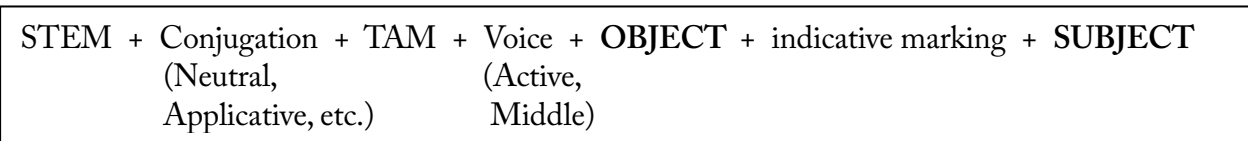   Applicative, etc.)   Middle)

Diagram 1: The structure of the Santali verb

For our discussion it is only important to note that object marking appears before the indicative marker *=a*, which appears before subject marking, cf. (22):

Santali

Transitive:	Intransitive:
(22) *dal=iɲ=a=e*	*hij-oʔ=a=e*
beat=1SG.OBJ=IND=3SG.SUBJ	come-MID=IND=3SG.SUBJ
'he will hit me'	'he will come'

However, a limited number of lexemes take an "object" but no "subject" (in purely structural terms), especially predicates involving an experiencer. Several other Munda languages have a similar class, such as Mundari (cf. Osada 1992: 104-109) and Gorum (cf. Anderson and Rau 2008: 398, 406).

Santali
(23) *rɛŋɛc=iɲ=a*
    hungry=1SG.OBJ=IND
    'I'm hungry'

---

[23] Neukom (2001: 122) refers to this unit as the "indirect object". However, as it is the primary function of the applicative to derive an object from a non-object, and only one object may be marked on the predicate, I merely assume the presence of "objects" in Santali.

This pattern is relevant to our discussion because it is also found in some portions of the copular paradigm in North Munda. This is easily explained since, e.g., Santali *mena* 'exist', which belongs to the same class as *rɛŋɛc* in (23), also functions as the non-negative existential/locative copula. Table 6 (from Neukom 2001: 168) provides an overview of the non-negative existential/locative copula in Santali.

	Singular	Dual	Plural
1 INCL	*mena=ɲ=a*	*mena–ʔ=laŋ=a*	*mena–ʔ=bon=a*
1 EXCL		*mena–ʔ=liɲ=a*	*mena–ʔ=le=a*
2	*mena=m=a*	*mena–ʔ=ben=a*	*mena–ʔ=pe=a*
3 ANIMATE	*mena=e=a*	*mena–ʔ=kin=a*	*mena–ʔ=ko=a*
3 INANIMATE	*mena–ʔ=a*		

Table 6: The Santali existential/locative copula (non-negated)

In Table 6 the glottal stop appears at first glance to be some kind of suffix which immediately follows the stem in the 3rd person, singular, inanimate as well as in the dual and plural (shaded grey in Table 6). This immediately raises the question as to the status of /ʔ/ here, as it is not found throughout the entire paradigm but only in some persons.

A look at the two closely related languages Mundari and Ho is helpful: In these two languages the glottal stop has further spread to ALL persons, including the first and second persons, singular, so that /ʔ/ has now become part of the stem of the copula in these two languages (cf. Deeney 1975: 47f. and Osada 1992: 118, respectively). This suggests that /ʔ/ in all three languages has spread throughout the respective paradigm via analogical leveling from some initial position where it once fulfilled a particular function, while in Santali these three persons in the singular (all animate) have resisted this change.

I assume that /ʔ/ in all three paradigms derives from the genitive / nominalizer =*aʔ*, i.e., *mena* + *aʔ* ⇒ *menaʔ*, and that this form originally served to mark the 3rd person, singular, inanimate and then spread via analogical leveling to the dual and plural, as languages often generalize one stem within a paradigm. Although this may seem somewhat *ad hoc* at first glance, there are convincing reasons for assuming such a development. To begin with, as we saw in Section 4.3, in many languages of the region a form which is homophonous with the genitive serves to mark the 3rd person, singular. Furthermore, and crucially, we noted in Section 4.3 that /ʔ/ regularly marks a 3rd person, singular, inanimate object in Santali in the imperfective and non-past applicative[24] and is also regularly found with 3rd-person subjects in Mundari and Ho. Taken together, this strongly suggests that this was the starting point of the "marker" /ʔ/ in the copula.

This analysis receives still further support from the suppletive negative form of this copula, shown in Table 7. The stem of this form appears to have been borrowed from IA, cf. Sadri *bʌn–*

---

[24] Neukom (2001: 168) also notes that this form is otherwise used to mark inanimate objects, although noting that its function in the non-singular copula is not clear.

'become'.[25] But whether *bən* is of Munda or IA origin, it is followed by the marker of the middle voice *-ok'* (via vowel harmony > *-uk'*). Hence, the glottal stop has a different status with the negated copula than with the non-negated copula and is not connected to any "nominalizer". Consider however the 3rd person, singular, animate (shaded grey); here we find the "nominalizer" *=icʔ* for animate beings instead of the usual animate marker of the 3rd person, singular, *-e* (Neukom 2001: 170, cf. also example (19) above).

	Singular	Dual	Plural
1 INCL	*bən-ug=iɲ=a*	*bən-uʔ=laŋ=a*	*bən-uʔ=bon=a*
1 EXCL		*bən-uʔ=liɲ=a*	*bən-uʔ=le=a*
2	*bən-uʔ=me=a*	*bən-uʔ=ben=a*	*bən-uʔ=pe=a*
3 ANIMATE	*bən-ug=icʔ=a*	*bən-uʔ=kin=a*	*bən-uʔ=ko=a*
3 INANIMATE	*bən-uʔ=a*		

Table 7: The Santali existential/locative copula (negated)

The presence of the animate nominalizer *=icʔ* in the form *bən-ug=icʔ=a* [NEG.COP-MID=NMLZR:ANIM=IND] 's/he is not' by analogy also suggests that the /ʔ/ found in the non-negated paradigm derives from the genitive/nominalizer, which then spread via analogical leveling: As *=icʔ* to my knowledge fulfills no other function in Santali, it clearly serves as a nominalizer here to mark the 3rd person, singular, animate. All this suggests that at some earlier time, nominalizers were used productively to denote the 3rd person, singular with the copula and a number of other forms.

In sum, I believe that in both the non-negated paradigm (Table 6) and in the negated paradigm (Table 7) the nominalizer first appeared in the 3rd person, singular instead of the usual person marker – for reasons which are not yet entirely understood – and then gradually spread to other parts of the non-negated paradigm by analogical leveling. Only the 1st and 2nd persons, singular, and the 3rd person, singular, animate in the non-negated paradigm in Santali have resisted this trend, whereas in Mundari and Ho this form has further spread to ALL persons, so that it must now be considered a part of the stem.

We now turn to the copular forms in Sadri, which show a number of parallels to the North Munda forms. First consider the four suppletive present-tense copulas in Sadri given in Tables 8 and 9, with the four-way negative/non-negative and identificational/existential-locative oppositions mentioned by Osada (1991) and discussed above in section 2.5.

---

[25] Perhaps motivated by the similarity to the native negative marker *baŋ* 'NEG'? Note also that the cognate form to Sadri *bʌn-* in neighboring Bengali, *bɔn-/bon-*, may also be interpreted as 'become', but with rather negative connotations, i.e., 'be reduced to' (cf., e.g., Radice 1994: 224 note 8).

Person	Affirmative		Negative	
	Singular	Plural	Singular	Plural
1	*he–k–ō*	*he–k–i*	*nʌ–lag–ō*	*nʌ–lag–i*
2	*he–k–is*	*he–k–a*	*nʌ–lag–is*	*nʌ–lag–a*
3	*he–k–e*	*he–k–ʌẽ*	*nʌ–lag–e*	*nʌ–lag–ʌẽ*

Table 8: The identificational copula in Sadri

Person	Affirmative		Negative	
	Singular	Plural	Singular	Plural
1	*ahō*	*ahi*	*nʌ–k–hō*	*nʌ–k–hi*
2	*ahis*	*aha*	*nʌ–k–his*	*nʌ–k–ha*
3	*ahe*	*ahʌẽ*	*nʌ–k–he*	*nʌ–k–hʌẽ*

Table 9: The existential / locative copula in Sadri

The forms given in Table 10 are found in periphrastic forms. These derive from the affirmative existential / locative copula, which came to function as an auxiliary and eroded phonologically. Note that some TAM categories take the form with /h/ while others take the form without it.

Person	Singular	Plural
1	*–(h)ō*	*–(h)i*
2	*–(h)is*	*–(h)a*
3	*–(h)e*	*–(h)ʌẽ*

Table 10: Personal marking deriving from the reduced copula in periphrastic forms

The "-*k*-less" copular forms apparently all derive from OIA: According to Tiwari (1960: 177, §592), *ah*- and the non-present stems *bhe*- and *ho*- (not given in the tables) derive from the OIA roots √*as* and √*bhū*, the details of which need not concern us here; *nʌ-lag*- has a different etymology – it derives from the negative marker *nʌ* and the verb *lag*- 'attach', which has also assumed a large number of other functions. The forms with -*k*, on the other hand, shaded grey in Tables 8 and 9, have traditionally been somewhat puzzling to researchers. Jordan-Horstmann (1969: 79, 94, fn. 3) notes that this element, as with the -*k* of the infinitive, may derive from the OIA diminutive marker -*k*, although she also calls this into question, as I do as well (cf. section 4.2 above).[26]

---

[26] Tiwari (1960: 178) lists similar forms in "Standard Bhojpuri", *ho/hokh* (affirmative) and *naikhe* (negative), and notes that *hokh* is especially difficult to account for etymologically. Shukla (1981), on the other hand, describing the northern-most varieties of Bhojpuri, cites no forms of the copula containing a /k/, which is thus restricted to the more southerly varieties of Bhojpuri and to Sadri.

I believe that *–k* derives from the infinitive marker, which in turn derives from the genitive. This perhaps came about through a construction which consisted of the infinitive followed by the copula-as-auxiliary. This development would have then followed a path of development something like the (purely hypothetical) development in (24).[27]

(24) **ho-ek ahe* [COPULA-INF AUX] > **hoek=he* > **hek=he* > *heke*.

It is not possible at present to determine the semantics of this "construction", if it ever in fact existed. There is one construction in modern Sadri in which an infinitive is followed by a copula to denote obligation, as in (25). However, it is highly unlikely that this was the source for the further development of the copula, due to both the semantics of this construction and to the fact that *–k* would then only have been retained from this construction in two copular forms and with no other verb.

Sadri

(25) *aij*	*rait=ke*	*ja-ek*	*ho-w-i*	*ki*	*ni*	*ja-ek*	*ho-w-i?*
today	night=OBL	go-INF	NPRS.COP-*w*-FUT.3SG	or	NEG	go-INF	NPRS.COP-*w*-FUT.3SG

'Should [we] go there tonight or not?'   (Jordan-Horstmann 1969: 121)

It is likely that the copular paradigm shown in Tables 6 and 7, which is typical for North Munda in general, served as the model for the Sadri paradigm. This development took place in Sadri with strictly language-internal means, i.e., with the infinitive marker *-ek*, itself deriving in part from the "genitive marker" *=(ʌ)k* (section 4.2). However, a number of facts suggest that this structure was "imported" into Sadri from North Munda, above all the fact that Sadri has no object marking on the predicate, so that this position on the predicate in Sadri, where *–k* is found, does not "really" exist. Furthermore, although it is possible to explain the *–k*-form copulas as the reinterpretation of a periphrastic construction, there is no apparent Sadri-internal motivation for this, as this "construction" only applies to two forms of the copula but to NO OTHER VERB and does not fulfill any apparent function.

On the other hand, there IS motivation for these forms in the North Munda languages, where we have an internal object position, in addition to the external subject position, and a small, semantically motivated class of intransitives which mark for an object but not a subject. This small class also contains two of the four present-tense copular forms in one of which, the non-negated form, a nominalizer originally marked the 3rd person, singular, inanimate which then spread to most other persons in Santali and to all other persons in Mundari and Ho. Thus, although the Sadri forms *he-k-e* and *nʌ-k-he* are perhaps diachronically analyzable as periphrastic forms, their development appears to be entirely unmotivated unless we consider it against the backdrop of North Munda.

Although the discussion in this section is admittedly somewhat speculative, it is a fact that there is a /k/ ~ /ʔ/ directly following the stem (from an etymological perspective) of the copula in many languages of the region and that this element appears to be a relatively recent innovation in all languages in which it is found. Hence, even if the reader is not inclined to accept the

---

[27] Recall from the discussion above that /h/ is retained in some verbal categories but not in others. An account of this distribution awaits further study.

development suggested above, the data nevertheless point towards convergence between these two families with respect to these forms, whatever the original source, i.e., even if the genitive was not directly involved in the emergence of these copular forms, contact between speakers of these languages was clearly responsible for the spreading of this feature throughout the region.

4.5   From genitive to focus marker

A final function which the "genitive" marker has assumed in many languages of the region is that of marking focus. Consider the following example from Sadri, where the "genitive" serves as a focal marker, although Jordan-Horstmann (1969: 107) considers the morpheme in this case to be semantically empty ("∅").

Sadri
(26) *sob*   *bhai*   *kʌh-l-ʌẽ:*   *"hã,*   *cʌl-a!*   *aij*   *huā=e=kʌr*   *ja-b!"*
     all   brother   say-PST-3PL   yes   go-IMP.2PL   today   there=FOC=FOC[28]   go-FUT.1PL
     'All the brothers said: "Yes, let's go! Today we will go <u>there (i.e., nowhere else)</u>!"'

The status of this marker as denoting focus was confirmed by speakers I consulted, who noted that it highlights the place (*huā*), although they also unanimously considered its use sub-standard. This might explain why Jordan-Horstmann analyzed it as semantically empty, as speakers tend to reject such examples at first.

Although I do not have evidence of this in North Munda languages, this focusing function of the genitive is also attested for the South Munda language Kharia where once again, most speakers at first reject the form as incorrect, although its use is quite common.

Kharia
(27) ... *ginir,*   *amkul,*   *banari*   *ro*   *choʔɖɖ=aʔ*[29]   *beʔʈ=ɖom=aʔ*   *ɲimi*   *aw=ki*   *baroya.*
     G.   A.   B.   and   small=FOC   son=3POSS=GEN   name   COP=MID.PST   B.
     '... Ginir, Amkul, Banari and the name of his <u>youngest</u> son was Baroya.'

This focusing function of the "genitive/nominalizer" is perhaps related to the "genitive subjects" in Oriya (NIA, Neukom and Patnaik 2003: 59ff.), spoken in Orissa, to the south of Jharkhand. These "genitive subjects", all of which are complex nominals, trigger verb agreement just like nominative subjects and can freely be replaced by these, apparently without any semantic difference. In light of the discussion above, it is likely that this results from an original focus construction, although if so, any sense of focus seems to have been lost.

---

[28] The =*e* in this form is also a focal marker, and =*kʌr* would seem to "reinforce" this marker here. It is possible that =*kʌr*, whose exact function is still somewhat unclear, denotes contrastive focus, while =*e* is a more general "focus" marker. Further work with corpora is required here, especially with respect to =*kʌr*, due to its sub-standard status. However, as noted in the text following (26) above, all speakers agreed that the presence of =*kʌr* served to highlight the status of *huā* in this example.

[29] "NPs" in Kharia are marked only once for case, hence =*aʔ* on *choʔɖɖ=aʔ* in (27) cannot be interpreted as case agreement.

Oriya (adapted from Neukom and Patnaik 2003: 59)

(28) <u>*bhaghɔ=rɔ   singhɔ=rɔ*</u>   *sathi*   *bɔs–ith–il–e.*
    tiger="GEN" lion="GEN"   friend   sit-PERF-PST-3PL
    '<u>The tiger and the lion</u> made [became?, JP] good friends.'

       Here as well, questions as to the origin of this construction await further study, nevertheless, regardless of its origin, this represents yet another common development of the erstwhile genitive marker and again one which can easily be reconciled to the notion of nominalization, as nominalizations are often found in focus structures (cf., e.g., Bickel 1999).[30]

5   Summary and outlook

       It is well-known that the IA languages have had an enormous impact on all aspects of the Munda languages of central and eastern-central India. However, the Munda languages also appear to have been the catalyst for a number of changes in IA languages, at least when these are spoken in areas with a high concentration of speakers of Munda languages, such as Sadri, the traditional IA *lingua franca* for speakers of various Munda and Dravidian languages in much of Jharkhand.

       As I have shown in the preceding pages, a number of linguistic convergences can be found in the languages of Jharkhand, with some of these originating in Munda, others in IA, and others of unclear origin. While it is still much too early to generalize from these data in any detail, we can broadly summarize the results of this study as follows:

- Some of the features discussed here, such as the similarities between expressions for 'from' and 'to' and 'begin' and 'keep on', are true not only of the languages of Jharkhand but also of at least some languages further to the east, such as Bengali.
- Other features are true of only some of the languages of Jharkhand and of those further to the north, e.g., the use of the genitive to mark the 3rd person, singular, which is also found as far afield as Maithili, spoken in Bihar and Nepal.
- Yet others seem to set the languages of Jharkhand off from their immediate neighbors entirely, such as the grammatical distinction between alienable and inalienable possession.[31]
- Finally, features are also found which hold for some languages of Jharkhand and possibly also further to the south, such as Oriya (e.g., the use of the genitive to mark focus), or to the south and east, such as "anticipatory" markers or "conditional participles / converbs".

---

[30] Cf. also constructions such as that in (i) below, which shows that in Sadri as well, there is a close link between focus and other types of nominalization (note: *ja–* 'go' is here a so-called "vector verb" which follows the lexical stem marked by the linker *–e* and which denotes *Aktionsart*. Its exact semantics are irrelevant here):

Sadri

(i)   *je=kʌr* *se* *wiswas* *ho–e*   *ja–ela*   *ki*   *ʌb*   <u>*biha*   *ho–be*   *kʌr–i.*</u> [HKS.0019]
      CR=GEN ABL certainty COP-LNK go-PRS.3SG CMPL now wedding COP-NMLZR do-FUT.3SG
      'Through which certainty arises that the wedding <u>will definitely take place</u> (literally: "<u>the wedding will do being</u>").'

[31] This category is admittedly not uncommon in South Asia and is found, e.g., in IA languages further afield, as well as in many Munda and Tibeto-Burman languages. Nevertheless, it does not appear to be found in the IA languages directly bordering Jharkhand, which is noteworthy in this context.

Much further work is necessary before detailed hypotheses can be made with respect to earlier patterns of contact between the various ethnic groups of Jharkhand and their neighbors and whether Jharkhand, or perhaps Jharkhand, parts of Orissa and Bengal together, constitute a kind of "linguistic sub-area" of a larger "linguistic area" encompassing eastern and northeastern South Asia (cf., e.g., Moral 1998; Neukom 1999). Also, the extent to which Jharkhand and its neighbors to the north, extending up into eastern Nepal, form an area of convergence, a possibility suggested in Ebert (1993; 1999), also requires further examination. The present study is thus just one tiny stone in this mosaic, and further studies of this type, examining these and further features in an ever-increasing number of languages, will be necessary before we can even begin to reconstruct past patterns of contact between the many ethnic groups of this region. At any rate, as many other studies before this one have already demonstrated, the idea that all of South Asia is a single, more-or-less homogeneous *Sprachbund* is far too simplistic and requires further refinement.

The present study has been admittedly rather one-sided in its approach. First, for reasons of practicality, I have omitted Dravidian languages such as Kurukh and Malto, but these will eventually have to be taken into account for a more complete picture. The situation further to the south in southern Orissa, where many South Munda languages have long been spoken in direct contact with both IA and Dravidian, should also yield many interesting results when viewed from this perspective.

Finally, many of the languages in this region also possess normative literary standards which can diverge considerably from the varieties actually spoken in the region, e.g., Maithili, Bengali and Oriya, so that good normative grammars, even where these exist, will still leave many questions unanswered. Clearly, we have only just begun to scratch the surface in this highly complex but also linguistically fascinating region.

ABBREVIATIONS

1, 2, 3	person		LOC	locative
ABL	ablative		MID	middle voice
ACT	active		NEG	negation
ADD	additive focus		NIA	New Indo-Aryan
ANTIC	anticipatory telic		NMLZR	nominalizer
AUX	auxiliary		NPRS	non-present
CLASS	classifier		NPST	non-past
CMPL	complementizer		NSG	non-singular
COP	copula		OBJ	object
CR	correlative		OBL	oblique case
DU	dual		OIA	Old Indo-Aryan
EXT	extensional (case)		PASS	passive
FOC	focus		PERF	perfect
FUT	future		PL	plural
GEN	genitive		POSS	possessive
HUM	human		PRS	present

IA	Indo-Aryan		PST	past
IMP	imperative		PTCP	participle
IND	indicative		SEQ	sequential converb
INF	infinitive		SG	singular
INGR	ingressive		SUBJ	subject
INST	instrumental		TEL	telicizer
IPFV	imperfective		TOP	topic
L1	first/native language		TR	transitive
L2	second language		VOC	vocative
LNK	linker (derives from an earlier converb)		*w, y*	hiatus-breaking glides

REFERENCES

Abbi, Anvita. 1997. "Languages in Contact in Jharkhand." In: Abbi, Anvita (ed.), *Languages of tribal and indigenous peoples of India. The Ethnic Space*, 131-148. Delhi: Motilal Banarsidass.

Anderson, Gregory D.S. 2007. *The Munda verb: Typological perspectives.* Berlin, New York: Mouton de Gruyter [Trends in Linguistics, Studies and Monographs 174].

Anderson, Gregory D.S.; and Rau, Felix. 2008. "Gorum". In: Anderson, Gregory D.S. (ed.), *The Munda languages*, 381-433. London, New York: Routledge [Routledge Language Family Series].

Bickel, Balthasar. 1999. "Nominalization and focus constructions in some Kiranti languages". In: Yadava, Yogendra P.; and Glover, Warren W. (eds.), *Topics in Nepalese linguistics,* 271-296. Kathmandu: Royal Nepal Academy.

Chatterji, Suniti Kumar. 1926. *The origin and development of the Bengali language. With a Foreword by Sir George Abraham Grierson.* [Third impression: New Delhi: Rupa, 2002].

Coulson, Michael. 1976. *Sanskrit* (Revised by Richard Gombrich and James Benson). London: Hodder Headline [Teach Yourself].

Deeney, John J. 1975. *Ho grammar and vocabulary.* Chaibasa: Xavier Ho Publications, St. Xavier's High School.

Ebert, Karen H. 1993. "Kiranti subordination in the South Asian areal context". In: Ebert, Karen H. (ed.), *Studies in cause linkage: Papers from the First Köln–Zürich Workshop*, 83-110. Zürich: ASAS-Verlag.

Ebert, Karen H. 1999. "Nonfinite verbs in Kiranti languages – an areal perspective". In: Yadava, Yogendra P.; and Glover, Warren W. (eds.), *Topics in Nepalese linguistics,* 371-400. Kathmandu: Royal Nepal Academy.

Ghosh, Arun. 2008. "Santali". In: Anderson, Gregory D.S. (ed.), *The Munda languages,* 11-98. London, New York: Routledge [Routledge Language Family Series].

Grierson, Sir George Abraham. 1903. *Linguistic survey of India.* Vol. V*: Indo-Aryan family, eastern group, Pt. II: Specimens of the Bihari and Oriya languages.* Calcutta: Office of the Superintendent of Government Printing, India.

Grignard, A. 1924. *A Grammar of the Oraon Language and Study in Oraon Idiom.* Calcutta: Catholic Orphan Press.

Hoffmann, Revd. John Baptist. 1905. *Mundari grammar and exercises,* Parts I and II. [Reprint: New Delhi: Gyan Publishing House, 2001].

Jhā, Subhadra. 1958. *The formation of the Maithilī language.* London: Luzac and Company.

Jordan-Horstmann, Monika. 1969. *Sadani: A Bhojpuri Dialect Spoken in Chotanagpur.* Wiesbaden: Otto Harrassowitz [Indologia Berolinensis 1].

Kullū, Paulus. 1992. *Tonme Kol–āṅgre* (The New Testament in Kharia). Ranchi: Catholic Press.

Lewis, M. Paul. (ed.). 2009. *Ethnologue: Languages of the world, sixteenth edition.* Dallas, Texas: SIL International. Online version: http://www.ethnologue.org/web.asp.

Moral, Dipankar. 1998. "North-East India as a linguistic area". *Mon–Khmer Studies* 27: 43-53.

Neukom, Lukas. 1999. "Phonological typology of northeast India". *Linguistics of the Tibeto–Burman Area* 22.2: 121-147.

Neukom, Lukas. 2001. *Santali.* München: Lincom Europa [Languages of the World / Materials 323].

Neukom, Lukas; and Patnaik, Manideepa. 2003. *A grammar of Oriya.* Zürich: Universität Zürich [Arbeiten des Seminars für Allgemeine Sprachwissenschaft 17].

Noonan, Michael. 1997. "Versatile nominalizations". In: Bybee, Joan; Haiman, John; and Thompson, Sandra A. (eds.), *Essays on language function and language type in honor of Talmy Givón,* 374-394. Amsterdam: John Benjamins.

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

Osada, Toshiki. 1991. "Linguistic convergence in the Chotanagpur area". In: Mullick, S. Bosu (ed.), *Cultural chotanagpur: Unity in diversity,* 99-119. New Delhi: Uppal Publishing House.

Osada, Toshiki. 1992. *A reference grammar of Mundari.* Tokyo: Tokyo University of Foreign Studies.

Osada, Toshiki. 2008. "Mundari". In: Anderson, Gregory D.S. (ed.), *The Munda languages,* 99-164. London, New York: Routledge [Routledge Language Family Series].

Peterson, John. 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: 391-441.

Peterson, John. 2006. *Kharia: A South Munda language* Vol. 1: *Grammatical analysis.* Unpublished Habilitationsschrift, Universität Osnabrück.

Peterson, John. 2008. "Kharia." In: Anderson, Gregory D.S. (ed.), *The Munda languages*, 434-507. London, New York: Routledge [Routledge Language Family Series].

Pinnow, Heinz-Jürgen. 1966. "A comparative study of the verb in the Munda languages". In: Zide, Norman H. (ed.), *Studies in comparative Austroasiatic linguistics*. London, The Hague, Paris: Mouton and Co [Indo-Iranian Monographs V].

Radice, William. 1994. *Bengali: A complete course for beginners.* London: Hoddler Headline. [Teach Yourself].

Shukla, Shaligram. 1981. *Bhojpuri grammar.* Washington, D.C.: Georgetown University Press.

Taylor, John R. 1996. *Possessives in English. An exploration in cognitive grammar.* Oxford: Clarendon.

Tiwari, Udai Narain. 1960. *The origin and development of Bhojpuri.* Calcutta: The Asiatic Society [The Asiatic Society, Monograph Series X].

Verma, Manindra K. 1991. "Exploring the parameters of agreement". *Language Sciences* 13.2: 125-143.

Wohlgemuth, Jan. 2009. *A Typology of verbal borrowings.* Berlin, New York: Mouton de Gruyter [Trends in Linguistics, Studies and Monographs 211].

Yadav, Ramawatar. 2003. "Maithili". In: Cardona, George; and Jain, Dhanesh (eds.), *The Indo-Aryan languages,* 477-497. London, New York: Routledge [Routledge Language Family Series].

John Peterson
peterson@uni-leipzig.de