

John Peterson*

From “finite” to “narrative” – The enclitic marker =*a* in Kherwarian (North Munda) and Sadri (Indo-Aryan)

DOI 10.1515/jsall-2015-0010

Abstract: This study deals with what we refer to here as the “narrative marker” =*a* in Sadri, the traditional Indo-Aryan *lingua franca* of western and central Jharkhand in eastern-central India. We argue that this marker in Sadri is the result of the reanalysis of an inherited focus marker which originally attached to the present tense in Sadri and that this reanalysis was influenced by speakers of Kherwarian languages using Sadri as a *lingua franca*, as the distribution of this marker in Sadri closely resembles that of the homophonous finite marker in Kherwarian. Finally, we show how this feature contributes to further define a micro-linguistic area centered on western-central Jharkhand.

Keywords: Munda/Indo-Aryan contact, finite, narrative, reanalysis

1 Introduction

South Asia has long been of interest to researchers of areal linguistics, as the subcontinent is home to languages from five families (as well as at least two linguistic isolates), and as cities, towns and villages in South Asia are often multilingual, so that many of those living in the subcontinent are accustomed to using more than one language in their daily lives from childhood on. As is to be expected in such a situation, this has often led to convergences between neighboring languages – regardless of their genetic affiliation – sometimes even to the extent that it is possible to translate longer passages morpheme-for-morpheme from one language into another. While most scholars would probably no longer subscribe to the idea that all of South Asia is a homogeneous *Sprachbund*, recent work is increasingly showing that there are a number of micro-linguistic areas in the subcontinent in which long-term, intensive language contact has led to convergences between the languages involved. Thus a more realistic picture of the linguistic landscape of the subcontinent would probably have several

*Corresponding author: John Peterson, Department of General Linguistics, University of Kiel, Germany. E-mail: jpeterson@isfas.uni-kiel.de

micro-linguistic areas, each with a center or centroid and with a periphery which gradually “fades” into that of other, neighboring micro-areas.

One of these micro-areas is centered on the western-central portion of the state of Jharkhand in eastern-central India, to which a number of recent studies are dedicated (e.g., Osada 1991; Abbi 1997; Peterson 2010, to appear). In the present study, we discuss a further shared trait which has so far escaped the attention of researchers, namely the presence of the marker =*a*; this marker is found in all languages of the Kherwarian branch of North Munda as a marker of finiteness and, as we argue here, also in the Indo-Aryan language Sadri, the traditional *lingua franca* of western and central Jharkhand. We argue that this marker in Sadri is the result of the reanalysis of the inherited form *-la*, generally believed to derive from an enclitic focus marker **=la* which attached to the earlier present tense in a number of Indo-Aryan languages, including Sadri. We further argue that through multilingual speakers of Kherwarian languages using Sadri as a *lingua franca*, this form has been reanalyzed in Sadri as *-l*, which now forms part of the present-tense marking, and the enclitic narrative marker =*a*, the latter having a similar distribution to the homophonous marker in Kherwarian.

The remainder of this article is structured as follows: After a brief discussion of the languages spoken in eastern-central South Asia in Section 2, Section 3 presents a broad overview of the Sadri verb system, with special reference to the narrative marker. In Section 4 we take a closer look at the distribution of the homophonous marker in Kherwarian, before examining its distribution and function in detail in Sadri in Section 5. Section 6 then places this feature into the larger, areal perspective. Finally, Section 7 briefly summarizes the main findings and gives an outlook for future studies.

2 Sadri and its neighbors – the languages of Jharkhand and beyond

The state of Jharkhand is home to languages from three families - Indo-Aryan (henceforth “IA”, Indo-European), Munda (Austro-Asiatic) and Dravidian. Figure 1 shows our area of study.

This study focuses on Sadri, an IA language spoken throughout much of western and central Jharkhand. Sadri has traditionally been used as a *lingua franca* for speakers of the Dravidian and Munda languages of the region, although this function is now gradually being forfeited to the closely related Hindi, especially in the cities and larger towns.



Figure 1: Jharkhand and its neighbors.¹

Sadri has an array of alternate names, above all *Sadani*, which is the term generally used outside of India to refer to this language; however I will refer to it here exclusively as *Sadri*, as this is the preferred term in the region itself.² According to the Ethnologue, Sadri is spoken by 3,291,180 people (Lewis et al. 2014). Its closest linguistic relative, Bhojpuri, of which Sadri has often been considered a dialect, is spoken primarily to the northwest of Jharkhand by 39,716,000 people (Lewis et al. 2014).

As the present study deals with the influence of the Kherwarian branch of North Munda on Sadri, we now provide a brief overview of the major Munda

¹ The original map, which has been modified here and in Figure 2 somewhat, is copyrighted by sa PlaneMad/Wikipedia and can be downloaded at the following site (last accessed: 17 August 2014): http://de.wikipedia.org/w/index.php?title=Datei:India_Jharkhand_locator_map.svg&filetimestamp=20081229064837. I am grateful to Arun Ghosh for granting permission to use these maps.

² 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 and Sradri (Lewis et al. 2014).

Although this language is generally referred to in linguistic studies as *Sadani*, I will not use this term here, as the term *Sadani* in the region itself refers to a number of closely related languages, such as Sadri, Panch Parganiya, Kurmali and Khortha, all of which are spoken by the “Sadani”, the Indo-Aryan-speaking groups of western and central Jharkhand.

languages of the region which are in close contact with Sadri.³ All figures for numbers of speakers are from Lewis et al. (2014) and refer to the total number of speakers of the respective language.

The ca. two dozen Munda languages, which form the western branch of the Austro-Asiatic phylum, belong to one of two groups, which are not mutually intelligible, namely North and South Munda:

- *North Munda*, which is of primary interest here, consists of Korku, spoken further afield in central India, and the Kherwarian group, to which the three largest Munda languages belong: **Santali**, with 6,228,500 speakers, **Mundari**, with 1,120,280 speakers, and **Ho**, with 1,040,000 speakers. A number of other, much smaller languages belonging to this group are also spoken in or near Jharkhand, such as **Asuri**, **Birhor**, **Korwa**, **Mahali** and **Turi**. As the Kherwarian languages are largely mutually intelligible and as these smaller languages await serious documentation, we concentrate here on the three largest languages of this group, on which there is considerable published data.
- The only *South Munda* language spoken in Jharkhand is **Kharia**, with an estimated 241,580 speakers. Further languages from this group include **Juang**, spoken in central Orissa, and a number of other languages spoken in southern Orissa and further to the south in Andhra Pradesh, such as **Sora**, **Gadaba**, **Gorum**, **Remo**, etc. As these languages are outside our area of study, we concentrate on Kharia from this group.

Figure 2 provides an overview of the areas where the larger languages mentioned above are spoken and also includes the major IA languages of the neighboring regions, such as Bhojpuri, Magahi, Maithili, Nepali, Bengali and Oriya. These languages are spoken over extensive areas: For example, through migrations mainly in the 19th century, Mundari, Santali, Ho, Kharia and Sadri are now also spoken in eastern Nepal and further to the east. Figure 2 thus provides only a very general indication of the core areas of the respective languages.

³ Unfortunately, the present study cannot go into any detail on the relationship between IA, Munda and Dravidian. Although a full study of the languages of this region must of course include the North Dravidian languages Kurukh and Malto, these appear to be relatively new to this region, unlike the IA and Munda languages, which appear to have been in contact with one another much longer. For this reason and also due to a lack of detailed studies on these two languages, the inclusion of the Dravidian languages must be left to future research.



Figure 2: The major languages of eastern-central South Asia and their relative positions (simplified).⁴

3 A brief overview of the Sadri verb system

Sadri is in many ways a typical eastern-central Indo-Aryan language: It has no morphological ergativity and the finite verb agrees with the subject (=S and A, in typological terms) with respect to person, number and honorific status (PERS/NUM/HON). In addition, the finite verb is marked for tense, aspect and mood and, in the present tense, optionally for the narrative marker =*a*.⁵

⁴ Cf. fn. 2.

⁵ The following overview of the Sadri verb system is largely based on Peterson and Kiran (2011). It differs from the system in Jordan-Horstmann (1969) in a number of ways, for our purposes primarily in that the two categories which Jordan-Horstmann (e.g., 1969: 83) terms the “indicative ~ optative, present ~ indefinite” and the “indicative indefinite” are viewed here as the “present indicative” and the “subjunctive”, respectively. This is in line with the analysis in Nowrangi (1956: 90–91, 97–98), who likewise treats these as the “present” and “subjunctive”, respectively.

The verb system of Sadri contains morphologically finite and non-finite forms. The non-finite forms include the infinitive, the sequential converb, the imperfective and past participles, and the so-called conditional participle (actually a converb). These differ from finite verbs in that they do not inflect for PERS/NUM/HON or tense/mood. The finite verb, on the other hand, obligatorily appears in one of three modal categories – the indicative, the subjunctive, or the imperative – and does inflect for PERS/NUM/HON. The indicative distinguishes three tenses: past, present and future,⁶ and the imperative distinguishes two: present and future, cf. *kha* ‘eat (now)!’ as opposed to *kha-b-e* [eat-FUT.IMP-2SG] ‘eat (later)!’, a distinction which is typical of much of eastern Indo-Aryan.⁷ There are also a number of periphrastic categories which express both tense and aspect (imperfective and perfect). For further details on these categories and their functions, cf. Peterson and Kiran (2011).

Of these categories, the present tense is unique in that it may also be marked for =*a*. Table 1 gives the (simplified) present-tense paradigm of the verb *kha*-‘eat’ to illustrate this distribution.⁸ From a purely structural perspective, /*a*/ is not obligatory except in the second-person singular, according to Nowrangi (1956: 90–91), perhaps due to phonotactic reasons, as the form **khaisl* is not permitted in Sadri.⁹

Table 1: The present-tense paradigm of *kha*- ‘eat’ in Sadri (simplified) (cf. Nowrangi 1956: 90–91).

Person	Singular	Plural
1	<i>khaon(a)</i>	<i>khail(a)</i>
2	<i>khaisla</i>	<i>khawal(a)</i>
3	<i>khael(a)</i>	<i>khaen(a)</i>

6 The future is in fact a combination of tense and mood, as it can denote either a definite future action or inference. For ease of discussion, we treat it here as a tense, as its exact status is not relevant for our discussion.

7 While the future imperative is homophonous with the future indicative, the two differ with respect to their negators and can thus be considered distinct categories, cf. *ni khabe* ‘you will not eat’ vs. *na khabe* ‘do not eat (later)!’.

8 The paradigms in Tables 1 and 2 (see below in main text) are simplified in that they only show the lowest levels of honorificity, i.e., the familiar unmarked 2nd persons and the unmarked 3rd persons. This is justified as the other honorific levels all derive from and are homophonous with other persons, e.g., the marking of the polite form of the 2nd person is homophonous with that of the 1st person plural, etc.

9 Although we have not yet been able to independently confirm that the 2nd person singular cannot appear without =*a*, e.g., that a form such as **khaisl* cannot be used, all occurrences of this form in our corpus are marked for =*a*.

The marker =*a* is enclitic: Its presence is not obligatory, as Table 1 shows (with the exception of the 2nd person singular). Furthermore, its scope is not necessarily limited to the verb to which it attaches, as example (1) shows: While the scope of =*a* in the first two predicates in (1), *mil-el* and *anen*, only extends to the verb to which it attaches, the final =*a* has scope over both the verbs *khan* and *pien* at the end of the sentence. (The scope of =*a* in each case is indicated by underlining.)

Sadri (IA)

- (1) *bec=le se paisa mil-el=a aur u=kar se hamar*
 sell=ANTIC then money be.found-PRS.3SG=NAR and that=GEN ABL 1PL.GEN
admi=man sag sabji cawal dail aur le=ke
 man=PL spinach vegetable unhusked.rice lentil.soup and take=SEQ
an-en=a *aur kha-n pi-en=a*
 bring-PRS.3PL=NAR and eat-PRS.3PL drink-PRS.3PL=NAR
 ‘When they sell it, they get money (=money **is found**), and from that our people
 get (=bring) sag, vegetables, rice and dal and [can] **eat [and] drink.**’
 [MB.017]

Due to this behavior, which is typical of enclitics but atypical of affixes, we consider =*a* in Sadri to be an enclitic, not a suffix. As we shall see in the following section, the homophonous form in Kherwarian is also enclitic.¹⁰

With respect to the etymology of this marker, Chatterji (1926: 997, §728), following Bloch (1919: 241, §242), derives /la/ from Old Indo-Aryan (OIA) *lā*¹¹ ‘take’. Tiwari (1960: 176, §588), in his grammar of Bhojpuri, further suggests OIA *lag-* ‘touch; get attached’ as a possible source of this marker, via the development **lagita-* ‘attached’ (past passive participle) > *laia* > Bhojpuri *-la*, which he assumes “began to be attached with the radical tense (old present) to indicate an emphatic present” (Tiwari 1960: 176, §588). As we will see in Section 5, the analysis of /la/ as a focus marker fits in well with the analysis we propose of =*a* in Sadri, while /l/ has now become part of the marking of the present tense.

¹⁰ There are of course many further aspects which should be dealt with in a full discussion of the status of =*a* in Sadri as enclitic, including phonological issues for which at present data are lacking, but what is at issue here is primarily the fact that =*a* is not a necessary part of the word, unlike e.g. *-(e)l* ‘PRS.3SG’ or *-(e)n* ‘PRS.3PL’, and that the scope of =*a* is not necessarily restricted to the verb to which it attaches. For a further discussion of the issues involved, cf. Aikhenvald (2002) and Anderson (2005).

¹¹ To avoid ambiguity we follow Indological tradition in portraying /a/ as <ā> and /ʌ/ as <a> in Old Indo-Aryan (OIA) and Middle Indo-Aryan (MIA), while the forms <a> and <ʌ> are retained here for New Indo-Aryan (NIA).

The “old present”, to which $*=la$ originally attached, has also evolved further without this marker to become the subjunctive in modern Sadri, a quite common development in IA in general (cf. Masica 1991: 273, 281–282).¹² Table 2 provides a paradigm of the present-day subjunctive in Sadri, the morphological continuation of the “old present”.

Table 2: The subjunctive paradigm of *kha-* ‘eat’ in Sadri (simplified) (cf. Nowrangi 1956: 97).

Person	Singular	Plural
1	<i>khaon/khaō</i>	<i>khai</i>
2	<i>khais</i>	<i>khawa</i>
3	<i>khæ</i>	<i>khaē</i>

We follow Jordan-Horstmann (1969: 79) in assuming that the alternative form *-l* of the (modern) present-tense marker is the short form of *-la*, so that *-l* historically derives from *-la* through the omission of */a/*. There remain a number of questions however, such as the lack of an */l/* in the 1st person singular and the 3rd person plural (cf. once again Table 1), which can only be dealt with in passing here: Recalling that we assume that $*=la$ originally attached to the “old present”, i.e. the modern subjunctive (Table 2), this means that in all persons other than these two, */la/* followed either an oral vowel (*khæ*, *khai*, *khawa*) or */s/* (*khais*). It is only in the 1st person singular and in the 3rd person plural that this form would have attached to an */n/* or a nasalized vowel – cf. *khaon/khaō* and *khaē*, respectively – and it is only here that the */l/* is lacking. We thus assume that in these two positions the sequence $*-n-la$ or $*\tilde{v}-la$ (where \tilde{v} represents a nasalized vowel) evolved into *-na*, later *-n(=a)*.¹³

Finally, although we share the view that *-l* is the reduced form of *-la* historically, in light of the presence of the homophonous form *=a* in the Kherwarian languages with a similar distribution, we assume that the development from *-la* to *-l* resulted through the reanalysis of *-la* at some point in time as *-l*, now part of the present marker, and *=a*, whose exact function we discuss in the following, i.e., */a/* must have been reinterpreted as a marker in its own right, similar to the homophonous form in Kherwarian. This development appears to be unique to Sadri within the IA languages and is to my knowledge found in no other IA language of the region. Otherwise, *=a* is only found in Kherwarian,

¹² What I term “subjunctive”, in line with the terminology in Nowrangi (1956), corresponds to Masica’s (1991) “contingent future”.

¹³ Cf. Bloch (1919: 239, §240) for a similar proposed development in Marathi.

hence we assume that the above-mentioned reanalysis of *-la* as “*-l + =a*” was originally motivated in Sadri by speakers of Kherwarian languages using Sadri as a *lingua franca*.

4 *=a* in Kherwarian

All of the North Munda languages of Jharkhand belong to the Kherwarian group and have an enclitic marker *=a* which attaches to the (clause-final) predicate in finite clauses, and we assume that it is from these languages – above all from Mundari and Ho – that the reanalysis of *-la* as *-l + =a* was motivated in Sadri. As the distribution and function of this form in Kherwarian are well understood, we begin with a closer look at this form in Kherwarian, which we can then use as our standard of comparison for the function(s) of the homophonous marker in Sadri.

=a is referred to by researchers of Kherwarian languages by different terms. For example, Hoffmann (1905/1909 [2010]) considers this unit to be a form of the “copula” in Mundari, which is also the analysis which Ramaswami (1992) appears to assume for the cognate form in Bhumij, which he glosses as “cop”, although I can find no further discussion of it in his grammar. Although this may be the original function of this form, a topic which we will not pursue further in this study, this is certainly not the function of this form in the modern languages, including Mundari and Bhumij.¹⁴ Several other terms are also found: E.g., writing on earlier studies, Pinnow (1966: 171) refers to this form as the “categorical” marker. More recent terms for this element, and in our view more appropriate with respect to the actual function of this marker, include “predicator” (cf. Osada 1992), “indicative” (e.g., Deeney 1975; Neukom 2001; Osada 2008), “finitizer” (Anderson 2007) or “finite” (e.g. Sinha 1975; Ghosh 1994 [2010], 2008; Pucilowski 2013), and as early as 1966 Pinnow notes that the “categorical” *a* “might also be called finite *a*” (Pinnow 1966: 172). Hence, this last term is the term we use in this study.

The distribution of this form in the various Kherwarian languages appears to be quite uniform, at least based on the available descriptions of these languages. Hence the distribution of this marker in Santali, for which this distribution is best described, can serve as the basis for our discussion of this marker in Kherwarian in general.

¹⁴ We note here in passing one proposed etymology for this marker in the Kherwarian languages: Pinnow (1966: 172) suggests that this marker was originally “a demonstrative particle meaning ‘that, that is’ or something to that effect”, noting that it appears to be a rather recent formation in Kherwarian which “is still in a formative stage.” Also, Anderson (2007: 132–133) suggests that this form may be cognate with the Korku finitizer *-bà*.

Figure 3 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 Kherwarian languages. For most intransitive predicates, the structure is the same except that there is no object marker.¹⁵

STEM + Conjugation + TAM + Voice + **OBJECT** + FINITE MARKER = *a* + **SUBJECT**
 (Neutral, (Active,
 Applicative, etc.) Middle)

Figure 3: The structure of the Santali predicate.
 Source: adapted from Neukom (2001: 61).

With respect to the position of the subject enclitic, the picture is not quite as simple as Figure 3 suggests. Simplifying somewhat we can say that the enclitic subject marker (for S or A) attaches either to the last element preceding the clause-final predicate or, if the predicate is the only word in the clause, to the predicate itself. On the other hand, the marking of the object (P) is always predicate-internal, directly following TAM and voice marking and preceding the finite marker (if present). Example (2) illustrates the marking of A and P on the predicate, and (3) shows that S is marked similarly to A.

Santali (Kherwarian)

(2) *dal-∅ = ɨn=a=e*
 beat-NPST:ACT=1SG.OBJ=FIN=3SG.SBJ
 ‘he will hit me’

(3) *hij-∅-oʔ=a=e*
 come-NPST-MID=FIN=3SG.SBJ
 ‘he will come’

Example (4) illustrates that the subject marker often attaches to the last element preceding the predicate as an anticipatory enclitic if the sentence contains other elements appearing before the predicate, e.g., the subject (as in [4]), the object or an adjunct.

¹⁵ A limited number of lexemes in many Munda languages take an “object” but no “subject” (in purely structural terms), above all predicates involving an experiencer (cf. Neukom 2001: 116 on Santali; Osada 1992: 104–109 on Mundari; Anderson and Rau 2008: 398, 406 on Gorum). As these are not central to the discussion at hand, they will not be dealt with further here.

Santali (Kherwarian)

- (4) *khangə jhɔ̃ɛ=dɔ=e dər-ket'=a*.¹⁶
 then Jhore=TOP=3SG.SBJ run-PST:ACT=FIN
 ‘Then Jhore ran away.’

(adapted from Neukom 2001: 113)

With respect to the distribution of =*a* in Santali, Neukom (2001: 145) writes that it “appears only with lexemes in finite function, i.e., where the lexeme performs the function of a main predicate” which is restricted to declarative and interrogative sentences (but also the negative imperative). Neukom (2001: 145–146) explicitly writes that it is not found in the following environments:

- in subordinate clauses
- in imperatives
- where pragmatic enclitics such as the topic marker =*dɔ*, the focus marker =*ge* or the modal particle =*ma* are attached to the predicate
- sentences marked by a modal particle such as *cɔŋ* ‘ever, possibly’, the modal particle *ma*, and *paset* ‘perhaps’, when the predicate is within its scope;
- “in interrogative sentences suggesting a certain answer” (Neukom 2001: 146), where the rhetorical character is often reinforced by the negative particle *baŋ*. This is shown in examples (5) and (6), both adapted from Neukom (2001: 146). Here, as in the following, the symbol = \emptyset marks where the finite marker =*a* could have been added but was not:

Santali (Kherwarian)

- (5) *baŋ=dɔ turtə=dɔ=e si-o?kan=∅? iŋ=ge cɔ=ŋ*
 NEG=TOP Turta=TOP=3SG.SBJ plough-MID-IPFV 1SG=FOC ever=1SG.SBJ
si-o?-kan-tahẽkan=a
 plough-MID-IPFV-COP:PST=FIN
 ‘Was it really Turta who was ploughing? (No,) it was me who was ploughing.’

¹⁶ I have adapted the data from other studies here by substituting the sign “=” for “-” to indicate that I consider a certain morph to be enclitic rather than a suffix, which is often – although not always – in line with other researchers’ analyses, despite their use of the sign “-”. Further minor changes with respect to the gloss and transliteration have also been silently made here for the sake of uniformity.

- (6) *un=əʔ=pe* *daya-w-ad-e=reaʔ* *gun=do*
 that.much=NMLZ:INAN=2PL.SBJ feel.pity-W-APPL:PST:ACT-3SG.OBJ=GEN gratitude=TOP
okor-ta-e=∅
 where-POSS-3SG.SBJ
 ‘**Where was his**’ (i.e. the leopard’s) gratitude for your pity towards him?’
 (suggested answer is: Nowhere!)
 (adapted from Neukom 2001: 146)

As we shall see in the following section, while not identical, the distribution of =*a* in Sadri nevertheless closely resembles that of Kherwarian.

5 The distribution of =*a* in Sadri

In this section we examine the distribution of =*a* in Sadri based on its occurrence in spoken texts collected by this author. As we saw in Section 3 above, =*a* in Sadri – like the homophonous =*a* in Kherwarian – is a non-obligatory enclitic marker. What remains to be shown is to what extent the distribution and function of this marker resemble those in Kherwarian, as discussed in Section 4.

In our corpus of Sadri, which currently contains 7,534 words, there are altogether 116 present-tense verb forms which could host =*a*, of which 62 (53.4%) are marked by =*a* while 54 (46.6%) are not. This corpus consists of two main groups, differing primarily in age, but also to some extent in genre: 1) three texts from Jordan-Horstmann (1969) and one short story from Nowrangi (1956) and 2) narratives collected by this author during a research trip to Jharkhand in 2009.¹⁷ These two groups differ somewhat in their distribution in that the published texts, which are now all ca. 50 years old, contain only 35 present-tense forms, of which 16 (45.7%) are marked by =*a* while 19 (54.3%) are not, while the narratives have 81 present-tense forms capable of hosting =*a*, of which 46 (56.8%) host =*a* and 35 (43.2%) do not. This is summarized in Table 3.

In the following discussion, we restrict ourselves to the distribution of =*a* in the narratives which this author collected with native speakers in Jharkhand. This restriction is due to several reasons, the primary reason being that Sadri appears to have changed since the texts given in Jordan-Horstmann (1969) or Nowrangi

¹⁷ The older texts are ‘The brothers and their sister’ (Jordan-Horstmann 1969: 106–119), from a text collection by P.S. Nowrangi, which I have not been able to locate, ‘The jackal and the camel’ (Jordan-Horstmann 1969: 119–126), recorded by Heinz-Jürgen Pinnow in 1956, and ‘Budhu, the cook’ (Jordan-Horstmann 1969: 126–141), also from the text collection by P.S. Nowrangi (cf. Jordan-Horstmann 1969: 16–17). The corpus also includes the text ‘The divinating Muslim’, which I have taken from Nowrangi (1956: 160–166).

Table 3: The distribution of =*a* according to text groups in this author’s corpus.

Texts	Total present-tense forms	Present-tense forms that host = <i>a</i>		Present-tense forms that do not host = <i>a</i>	
		Number	Percentage of total	Number	Percentage of total
All texts	116	62	53.4%	54	46.6%
Older, published texts	35	16	45.7%	19	54.3%
More recent narratives	81	46	56.8%	35	43.2%

(1956) were composed. This is evidenced among other things by the fact that there are a number of forms in the texts from Jordan-Horstmann (1969) which we analyze as subjunctive, due to their function in the modern language, but which Jordan-Horstmann considers to be indicative forms which are underspecified (“indefinite”) with respect to tense and aspect, in line with their function in her texts. Crucially, modern speakers rejected examples such as *kɪhe* in (7) and accepted only those forms which we consider here to be “present tense” (cf. again Table 1).

Sadri (IA)

- (7) *sɪrdar siyar kɪh-e* “*dekh-a ...*” (Cf. modern *kɪhel(a)*)
 chief jackal say-“SBJV”.3SG look-IMP.2PL
 ‘The chief jackal says: “Look! ...”’
 (Jordan-Horstmann 1969: 127)

But even when we restrict our attention to the recent texts, we still find considerable differences with respect to the occurrence of =*a* in the individual narratives. Table 4 summarizes this information for the various texts, listed alphabetically according to the abbreviation of the text name (“Text” in Table 4).

Table 4: The distribution of =*a* in the more recent narratives.

Text	Total number of present-tense forms	Number of present-tense forms marked by = <i>a</i>	Percentage of present-tense forms marked by = <i>a</i>
BCB	20	16	80%
DD	3	0	0%
HKS	18	18	100%
MB	13	12 ¹⁸	92%
SBO	8	1	12.5%
SBT	19	0	0%

¹⁸ Actually, two verbs in this text in one case share the same finite marker =*a*; cf. *khan piena* in example (1). Since the scope of =*a* here extends to both verbs, we count both as being marked by =*a*.

We begin our discussion with text MB, in which all present-tense forms are marked by =*a* except for one. This unmarked form is the predicate of a subordinate clause, a relative clause, while all forms marked by =*a* are the predicate of a matrix or “finite” clause. The “=*a*-less” form is given in (8) (subordinate clause is underlined). For an example of four of the forms marked by =*a* in a matrix clause in this narrative, cf. once again example (1).

Sadri (IA)

- (8) ... *aur rʌŋ birʌŋ=kʌr* *carʌi=mʌn=bhi aɦʌē*¹⁹
 and of.different.COLORS=GEN bird=PL=ADD COP.PRS.3PL
je=ke ɦʌmre ciŋiʌ ***kʌɦ-il=∅*** *Hindi mē*.
 CREL=OBL 1PL “ciŋiʌ” say-PRS.1PL Hindi LOC
 ‘... and there are also birds of all different colors which we call *ciŋiʌ* in Hindi.’
 [MB.021]

This closely parallels the Kherwarian data, where “finite” forms, i.e., predicates in main clauses, are marked by =*a*, while subordinated forms are not. We now turn to another text, SBO, which in a sense is the opposite of MB, as here only one present-tense form is marked by =*a* while the remaining present-tense forms are not. In this narrative the speaker is telling the folk-history of his ethnic group/caste,²⁰ the weavers, and how they came to be referred to as Chik Baraik. The one present-tense form which is followed by =*a* in this story, a matrix or “finite” clause, is given in example (9).

Sadri (IA)

- (9) *ɦamre Cik Bʌŋaik ʌŋjait* *mē* ***a-w-il=a***.
 1PL Chik Baraik ethnic.group LOC come-W-PRS.1PL=NAR
Cik Bʌŋaik ek ... *Jɦarkɦŋd mē*
 Chik Baraik one Jɦarkhand LOC
rʌɦ-ek=wala battis ʌŋjait *mē se ek=ɦo heke*.
 live-INF=ADJZ thirty-two ethnic.group among one=CLF COP.PRS.3SG
 ‘We belong to (=come in) the Chik Baraik ethnic group. The Chik Baraik are one... one of the 32 ethnic groups living in Jɦarkhand.’
 [SBO.3-4]

¹⁹ Present-tense copulas cannot appear with =*a*, hence there is no =∅ following *aɦʌē* in (8), *ɦeke* in (9), *aɦe* in (10), etc.

²⁰ Speakers freely alternate between the terms *ʌŋjait* ‘ethnic group’ and *jati/jait* ‘caste’, so that these English terms should not be taken too literally.

With respect to the remaining present-tense forms, none of which is marked by =a, all but one fall into one of the two now familiar groups:

- With two exceptions, all of the present-tense forms which are not marked by =a are the predicate of a subordinated clause, as with the relative clause in example (10) or the object clause in example (11) (subordinate clauses are underlined).

Sadri (IA)

- (10) *hamre=ke se=kar piche ek=tho kahani ahe ki*
 1PL=OBL that=GEN behind one=CLF story COP.PRS.3SG COMP
ek dhāw gāw mē Jharkhand
 one time village LOC Jharkhand
mē ek=tho prātiyogita je kah-āl ja-el=∅
 LOC one=CLF competition CREL say-PTCP PASS-PRS.3SG
aijkail usān ... ho-e rehe.
 nowadays such COP-LNK COP.PST.3SG

‘We have a story behind that (= to us there is one story behind that) in which once, in a village in Jharkhand, such a *prātiyogita* (=competition), which it is called nowadays, was taking place.’

[SBO.6]

- (11) ... *se=kar mē saub=ke kah-āl ge-lāk ki*
 that=GEN LOC all=OBL say-PTCP PASS-PST.3SG COMP
“*kehār ka kā<i>r sāk-āl=∅ dekh-a-w-a!*”
 who what do<LNK> be.able-PRS.2PL see-CAUS-W-IMP.2PL
 ‘... in which all were told “Show [us] who of you can do what!”’
 [SBO.7]

- In one further case, the finite marker is lacking in a rhetorical question. Cf. example (12).

Sadri (IA)

- (12) *Cik Baraik kale kah-āl ja-el=∅?*
 Chik Baraik why say-PTCP PASS-PRS.3SG

‘Why **are** the Chik Baraik **called** [that]?’ (Intended meaning: “I suppose you’re asking yourself why we’re called Chik Baraik, so now I’ll tell you.”)

[SBO.5]

Again, this distribution closely resembles that in Kherwarian, where =a is not found in subordinate clauses nor in rhetorical questions.

There is however one example in SBO where =a would be expected if =a were a finite marker but where it is not found. Consider example (13), which is best considered a side comment, in this case a comment which provides the listener with background information.

Sadri (IA)

- (13) *hamar jait=wala=man luga bin-ek=wala p̄ihla r̄ah̄ē aur*
 1PL.GEN caste=ADJZ=PL cloth weave-INF=ADJZ first COP.PST.3PL and
luga=man=e ki p̄arda ...
 cloth=PL=FOC or cloth
“p̄arda” aijkail kah-al ja-el=∅ muda
 “parda” nowadays say-PTCP PASS-PRS.3SG but
age prakrit aur apbhramsa mē cik rehe
 previously Prakrit and Apabhramsa LOC cik COP.PST.3SG
 ‘Those in our caste used to weave clothes or cloth ... **today it is called**
parda but earlier, in Prakrit and Apabhramśa, it was [called] cik.’
 [SBO.9]

Example (13) shows that we cannot simply equate the function of =a in Sadri to that in Kherwarian, although the two have similar distributions. Based on these facts, I venture the following, still rather tentative hypothesis about the function of =a in Sadri:

Hypothesis: The use of =a with the present tense signals information which is considered by the speaker to be directly relevant to the narrative.

With that, =a in Sadri is primarily a discourse-pragmatic marker directly connected to the notion of relevance, not a marker of finiteness. Hence we gloss it as ‘NAR’ for ‘narrative marker’. According to this hypothesis, we expect to find clauses in a narrative which further the plot to have a predicate marked by =a. This will generally be true of “finite” clauses, which tend to be informative (statements) or to request information (interrogatives). However, we do not expect to find it in side comments, such as (13), nor in rhetorical questions such as (12), as the rhetorical question does not itself further the narrative but rather serves to indicate that the **following** information is especially important.

Example (14) illustrates this with the English translation, where the rhetorical question from (12) above is underlined.

- (14) ‘The Chik Baraik are one... one of the 32 ethnic groups living in Jharkhand. Why are the Chik Baraik called [that]? We have a story behind that in which once, in a village in Jharkhand, there was what nowadays is called a competition in which all were told “Show who of you can do what!” [SBO.4-7]

This also holds true in other cases in which a present-tense verb is the predicate of a morpho-syntactically non-subordinated clause but is not marked for =*a*, as in example (15). This example contains the present-tense form *kaḥel*, unmarked for =*a*; (15) is the only example in this text of the root *kaḥ-* ‘say’ in the present tense (out of 15 altogether) which is not marked for =*a*. According to our hypothesis, this should mean that it is not highly informative or does not further the narrative. We argue here that this is in fact the case, as it is the information which follows this verb, i.e., the quote itself, which is focused (in this case, contrastive focus), whereas this is not the case with the other present-tense forms of this verb.

Sadri (IA)

- (15) a. *aur qhela kaḥ-l-ak siyar=ke ki “ca-l-u*
 and clump.of.dirt say-PST-3SG jackal=OBL COMP go-IMP.2POL
bou, raur age qeg-u. moẽ pache qeg-mũ.”
 VOC 2POL first jump-IMP.2POL 1SG afterwards jump-FUT.1SG
 ‘And the clump of dirt said to the jackal “Go on, you jump first. I’ll jump later.”’
- b. *siyar kaḥ-el=∅ “nihĩ, sãgi, raur age qeg. moẽ pache*
 jackal say-PRS.3SG no friend 2POL first jump 1SG afterwards
qeg-mũ.”
 jump-FUT.1SG
 ‘The jackal **says** “No, friend, you jump first. I will jump later.”’
 [BCB:8-9]

Similarly, in the following example by the same speaker, given here with its context (in English translation), we see that in (16)c it is the information which follows *dekhen* ‘they see’ (unmarked for =*a*) which is in focus, i.e. the fact that there is a bundle of bread, and not *ceiḥke dekhen* ‘having climbed up, they see’ with its ‘tail-head’ linkage.

Sadri (IA)

- (16) a. ‘But the small child who had seen the bundle earlier, he could not climb.’
 b. ‘The bigger kids managed with difficulty, but the smaller one could not climb up.’
 c. *ce<i>ɾ=ke* *dekh-en=∅* *to huã roʈi=kar poʈom hae.*
 climb-LNK=SEQ see-PRS.3PL TOP there bread=GEN bundle COP.PRS.3SG
 ‘They (=the big kids), having climbed up, **see** that there is a bundle of bread there.’
 [BCB.45-47]

This also shows that there is a certain amount of speaker-specific freedom connected to this marker: It could of course be argued that the verb ‘say’ will always be “less important” than the information contained in the direct or indirect quotation, and yet this speaker chooses to mark these forms in general with =*a*. However, she leaves =*a* out when the quote is strongly focused, as with the contrastive focus in (15)b. Thus, although the level of “narrative relevance” may be left to the speaker’s discretion, the underlying principle still holds.

=*a* is also lacking in clauses containing information which is considered by the speaker to be known, as in example (17), from a description of rural life in Jharkhand. Here the speaker is telling me where his home village is located, although he knew that I was relatively familiar with the area, hence the translation in (17) begins with “As you know”. Example (17) is from the beginning of this narrative and merely provides the background for what follows.

Sadri (IA)

- (17) *hamar arʌn gãw Simɖega jila* *paɾ-el=∅*,...
 1PL.GEN REFL village Simdega district fall-PRS.3SG
 ‘[As you know,] our village is in (=falls **[within]**) Simdega district, ...’
 [SBT.01]

Finally, that the narrative marker in Sadri is a discourse-pragmatic marker and not a grammaticalized marker of finiteness is shown by example (18)c, where we find a subordinate clause (underlined) whose predicate is marked by =*a*.

Sadri (IA)

- (18) a. ‘In this caste there is one main custom to wash the feet,’
 b. ‘which is also current today.’

- c. *ʃakħan ke=u gotiya ke=kr=o ghār a-w-en=a*
 CREL.TEMP who=ADD guest who=GEN=ADD house come-W-PRS.3PL=NAR
 ‘When some guest **comes** to someone’s house,’
- d. *to u ghār=kar ʃanani loʃa thari le-i=ke*
 TOP that house=GEN wife small.pot metal.plate take-LNK=SEQ
gotiya=kar goʃ dho-w-en=a,
 guest=GEN foot wash-W-PRS.3PL=NAR
 ‘then the wife of the house brings a metal plate with a water vessel on it
 and **washes** the guest’s feet,’
- e. *u=kar bad=e halcal puch-al ja-el=a.*
 that=GEN after=LOC well.being ask-PTCP PASS-PRS.3SG=NAR
 ‘after that one asks about the [guest’s] well-being (= the well-being **is
 asked**).’
 [HKS.8-10]

Although (18)c is a subordinate clause, it is just as relevant to the narrative as the other clauses in (18), and in fact contains more “new” information than (18) d, since the washing of feet has already been mentioned in (18)a. Thus, (18)c contains new information which is just as relevant for the narrative as the information in the following clause.

Again, we stress that this marker, being a discourse-pragmatic marker, will undoubtedly be subject to speaker-specific preferences, and it may well be that another speaker would choose to leave out =a in (18)c. Nevertheless we argue that its underlying function is always the same, namely to mark information as directly relevant to the narrative.

Our analysis of =a in Sadri differs considerably from what is generally assumed for the homophonous form in Kherwarian, discussed in Section 4, and the question naturally arises whether the two morphemes are really connected in any meaningful way. We argue that they are and that the different function of this marker in Sadri results from a reanalysis of the underlying distribution of =a in Kherwarian (or perhaps in the Sadri of native speakers of Kherwarian languages). Consider e.g. example (19) from Santali, which shows a complex sentence containing a subordinated and a main clause. Santali is in many ways a typical predicate-final language in that the predicate of the main clause is virtually always the last element in the clause and subordinate clauses generally precede the main clause.

Santali

- (19) *oʔaʔ-re **sen-kate** uni haʔam=dɔ hɔpɔn-tɛt' koʔa=e*
 house-LOC go-SEQ that(AN) old.man=TOP son-3POSS boy=3SG.SBJ
met-a-e-kan=a...
 say-APPL-3SG.OBJ-IPFV=FIN
 'When [he] reached (=having reached) home, the old man **said to** his son,
 ...'
 (adapted from Neukom 2001: 185)

Here we see the most typical pattern for the presence/lack of finite marking in Kherwarian, with a subordinated clause, unmarked for =a (whose place is taken in (19) by *-kate* 'SEQ'), followed by a main clause, whose predicate is marked by the =a. Although we cannot go into detail here on information structure and text frequency, if we assume for the moment that there is at least a *tendency* in language in general for the main clause to more often directly further the narrative than the subordinate clause, what we essentially have in Santali (and Kherwarian) is a clause, unmarked for =a, which is tendentially not the main thrust of the narrative, while the main clause tendentially furthers the narrative. It would seem that it is this interpretation of the function of =a which has found its way into Sadri – motivated on the one hand by the status of =a in Kherwarian, but at the same time reinterpreted (perhaps primarily by non-Kherwarian speakers of Sadri) in a slightly different manner, although the outcome of both analyses is quite often identical.

In sum, although further work is necessary to determine the exact status of =a in Sadri, our study shows that its distribution is largely predictable and closely resembles that found in Kherwarian, despite its somewhat different function in Sadri. It therefore seems reasonable to assume that the form and function of =a in Sadri – even in the speech of Sadri native speakers, from whom all of the examples in this study derive – were originally motivated by speakers of Kherwarian languages using Sadri as a *lingua franca*. As we shall discuss in the next section, this is merely one of a number of linguistic features which have spread throughout this region and which set this region apart from those surrounding it.

6 The areal context

In two recent studies (Peterson 2010, to appear) a number of the linguistic features discussed in Abbi (1997) and Osada (1991) are summarized which are typical of many languages of eastern-central India, and attention is drawn to

further features which are typical of this region but which escaped the attention of earlier researchers. In this section, a few of these will be briefly summarized and the finite/narrative marker will be discussed against this backdrop.

6.1 Alienable vs. inalienable possession

Many languages of the region distinguish between inalienable possession (e.g., with body parts and kin-relations), indicated by enclitic marking for the possessor on the possessum, and alienable possession, in which the possessor appears in the genitive and precedes the possessum. Although this distinction is not at all uncommon from a cross-linguistic perspective, this is significant since the IA languages of this region otherwise do not express this semantic distinction in this way, although it is typical of virtually all Munda languages, both North and South Munda. Hence it may be assumed to derive from Munda in this region. Cf. (20)–(21).

Inalienable

Alienable

Santali (Kherwarian)

(20) <i>hɔpɔn=me</i>	<i>am=aʔ</i>	<i>oʔaʔ</i>
son=2SG	2SG=GEN	house
‘your son’	‘your house’	

(adapted from Neukom 2001: 32)

Sadri (IA)

(21) <i>bhɔɔji=har=man</i>	<i>u=man=ɔk jɔmin</i>
sister.in.law=3POSS=PL	3=PL=GEN land
‘his/her sisters-in-law’	‘their land’

6.2 ‘start’ vs. ‘keep on’

In many IA and Munda languages of the region, a single morpheme can have both an inceptive and a durative (‘keep on’) or general imperfective interpretation.²¹ For example, *laʔ* in Kharia, which derives from Sadri

²¹ This ambiguity does not depend on the Aktionsart of the predicate, at least in those languages for which I have been able to verify this.

lag- ‘begin’, can denote both the inception of an action or event (22) or general imperfectivity (23), the latter being its most common function, despite the etymology.

Kharia (South Munda)

- (22) *hobne=te=ga ubar kole? kunḍu? jal=te bajhe=kon*
 that.much=OBL=FOC two parrot child net=OBL get.trapped=SEQ
 “*tāy tāy*” *toro²q=na la?²=ki=kiyar.*
 “tay, tay!” CRY=INF “IPFV”=MID.PST=3DU
 ‘Meanwhile (=in that much), two baby parrots got caught in the net and began crying “Tay! Tay!”.’
 [BB, 2:33]

- (23) ... *khaṛiya lebu=ki pujaṣaṭh karay=na la?²=ki=may.* ...
 Kharia man=PL sacrifice do=INF IPFV=MID.PST=3PL
 ‘... the Kharia men used to perform sacrifices ...’
 [AK, 2:6]

This ambiguity is also found in Sadri, as (24) shows: Speakers confirmed that without the first sentence in (24) and without *lb* ‘now’, (24) could also have the meaning ‘The old man and the old woman kept on thinking.’

Sadri (IA)

- (24) *rait bhe-l-ak. lb buḍha buḍhiya bicar kar-ek*
 night become-PST-3SG now old.man old.woman thought do-INF
he-l-lē *ki...*
 “begin”-PST-3PL COMP
 ‘It became night. Now the old man and the old woman began thinking ...’
 (adapted from Jordan-Horstmann 1969: 129)

6.3 ‘from’ and ‘to’

In many IA and Munda languages of this region we find similarities between markers with an ablative and an allative meaning. I refer to this as the “extensional” (‘EXT’) function, as it refers to the distance extending from or to an event or location. Cf. the two interpretations of the Sadri postposition *le* in example (25).

Sadri (IA)

- (25) *se=khan le hɔmre=mɔn=ke cik bɔɔaik kɔh-ɔl*
 that=time EXT 1NSG=PL=OBL Chik Baraik say-PTCP
ja-t=he. aij le.
 PASS-IPFV=PRS.3SG today EXT
 ‘Since that time we are called “Chik Baraik”. Until today.’

[SBO.15]

In Munda languages of the region, although the ablative and allative are generally distinct, they often show formal similarities, cf. e.g. Kharia *tay* ‘from’ and the allative postposition *khoʔtay* ‘up to’ (< **khoʔ tay* [place_{ABL}] ‘from the place’).

6.4 Anticipatory categories

Many North and South Munda languages of the region possess an “anticipatory” predicative category denoting that one event is directly followed by another. Consider example (26) from Santali.

Santali (Kherwarian)

- (26) *jɛmɔn=e bɔlɔ ɡɔt'-len=a, tɛmɔn=ge=kin kiləp ɛsɛt*
 as=3SG.SBJ enter TEL-ANTIC:[MID]=FIN then=FOC=3DU.SBJ close close
ɡɔt'-ked-e=a.
 TEL-ACT.PST-3SG.OBJ=FIN
 ‘The moment he got in, they closed and shut him up.’
 (adapted from Neukom 2001: 80f.)

IA languages of the region such as Sadri have a similar category, generally referred to as the “conditional” participle/converb, although this form is by no means restricted to conditional clauses. Consider the forms marked by the anticipatory suffix *-le* (etymologically unrelated to the postposition *le* discussed in Section 6.3) in example (27).

Sadri (IA)

- (27) ... *jolha u=ke poʔra-le ɔur ɡhoʔa ɔpɔɔ bɔiʔh-le*
 Muslim that=OBL grab-ANTIC and horse above sit.down-ANTIC
rɔ<i>h ge-l-ɔk.
 remain<LNK> TEL-PST-3SG
 ‘... the Muslim grabbed it and sat down on the horse and remained [sitting there].’
 (Nowrangi 1956: 165)

6.5 The genitive

Morphemes which derive from the genitive (and which are often still homophonous with an allomorph of the genitive) are found in many languages of the region fulfilling various functions. Here we discuss two of these, the origins of which are still somewhat unclear, but which are found in languages of both IA and Munda stock and which therefore may be assumed to have spread through multilingual speakers from one language (family) to another, whatever direction the “borrowing” may have taken.

6.5.1 Person marking

In many languages of the region, a marker which is homophonous with the genitive is used to mark person – generally the 3rd person singular – in various predicative categories. For example, in Sadri $-(\lambda)k$, which is homophonous with an allomorph of the genitive, is used to mark the 3rd person singular in the indicative past, cf. *kha-l-ak* [eat-PST-3SG] ‘s/he ate’. Similar evidence is found in Kherwarian languages, e.g., Santali, where forms homophonous with an allomorph of the genitive, $=(a)ʔ$, mark the 3rd person singular inanimate object in the imperfect and non-past applicative (Neukom 2001: 121–122).

6.5.2 The copula

Another convergence involves the copula, which often contains a $/k/ \sim /ʔ/$ in its stem deriving from the genitive in the respective languages. For example, the stem of the non-negated existential/locative copula in Santali, *menaʔ* (in all persons except the 1st and 2nd singular, and the 3rd singular animate), which derives from *mena* ‘remain, stay’ (cf. Peterson to appear, Section 3, for details and further examples). Similarly, in Sadri we find the present-tense stems *hek-*, the identificational copula, and *nakh-*, the negated existential/locative copula, with a $/k/$ deriving from the genitive. Interestingly, in Sadri’s closest linguistic relative, Bhojpuri, spoken primarily to the northwest of Jharkhand, this $/k/$ is only found in more southerly dialects, including the standard dialect (cf. e.g., Tiwari 1960: xxxiv, §22; 178, §595–597), i.e., those dialects which even today are spoken in the regions closest to Jharkhand, but not in those further to the north, as it is not found in the dialect described in Shukla (1981).

6.6 The Jharkhandi micro-linguistic area

In Peterson (to appear) the above-mentioned linguistic features are represented as isoglosses on a very rough linguistic “map” to help visualize the distribution of these traits throughout this region. This is not without problems, as this type of map gives the impression that the various traits fulfill the same functions in each language and that it is only the form that varies from one language to another, although similar categories can indeed have quite distinct distributions/functions in the various languages. To cite one example: In Peterson (2010: 60–61; to appear, Section 3.1), it is noted that virtually all languages of this region possess enclitic classifiers, found above all in conjunction with numerals. However, beyond this one environment, their distributions can vary considerably, as e.g. many IA languages of the region such as Oriya and Sadri also allow these classifiers to directly attach to nouns to denote specificity (cf. Neukom and Patnaik 2003: 24–34), which is not possible in Munda languages such as Kharia. As such, an isogloss map indicating that Oriya, Sadri and Kharia all have “classifiers” would correctly indicate that all of these languages have a similar category which we can term “classifiers” but at the same time it would obscure the fact that these categories can differ considerably from one language to another and that they are arguably not the “same” category. At the same time, isoglosses provide a quick and informative overview of the distribution of select linguistic phenomena within a specified geographical area. Hence, in the following we will also make use of isoglosses, keeping these limitations in mind.

Figure 4 presents a schematic “map” of the linguistic traits discussed above, from Peterson (to appear), with the presence of the finite/narrative marker =*a* added here, the distribution of which is shaded on the map. Note that without this marker, Sadri otherwise forms a tight bond with Mundari, Ho and Kharia; Santali, unlike these four languages, does not show similarities between ‘start’ and ‘keep on’. At the same time Sadri also forms a tighter bond with Mundari, Ho and Santali, whereas Kharia differs in that it does not have person marking with /k/ ~ /ʔ/ on the predicate, unlike these other four languages. This affinity between Sadri and the three Kherwarian languages becomes even more pronounced when we take the presence of the finite/narrative marker =*a* into account, as Kharia does not have this marker. Thus Sadri and the Kherwarian languages Mundari, Ho and to a somewhat lesser extent Santali, form the centroid of a Jharkhandi micro-area, followed by the South Munda language Kharia, and finally by the remaining IA languages of the region.

Again, we emphasize that such an approach obscures the many differences between the linguistic criteria used and any combination of criteria will necessarily be rather arbitrary, since other linguistic criteria could just as easily have been chosen. Nevertheless, the picture which arises is one of Sadri sharing many traits

with North Munda and to a somewhat lesser extent with Kharia, which is not surprising when one recalls its role as the traditional *lingua franca* of much of the Munda-IA contact area.

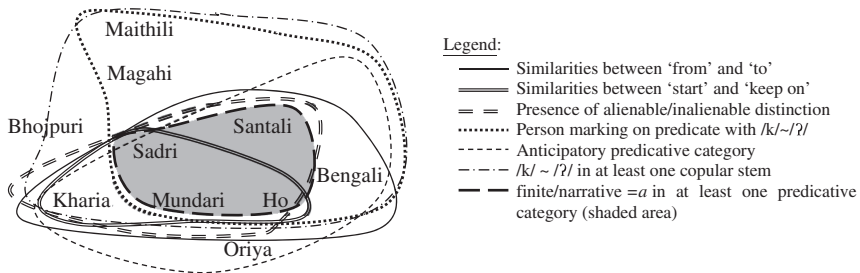


Figure 4: Convergence patterns of the investigated traits.

7 Summary and outlook

Beginning with Osada (1991) and Abbi (1997), a number of studies have begun calling attention to various convergences among the languages of Jharkhand, belonging to Indo-Aryan, Dravidian and Munda stock. As the preceding pages have shown, we can now add yet another probable candidate to this list, namely the narrative marker =a in Sadri. As we argue here, this marker is in all likelihood the result of the reanalysis of an earlier focus marker **=la*, of uncertain origin, which occurred in combination with the present tense in many Indo-Aryan languages, most likely to denote an “emphatic present” (Tiwari 1960: 176, §588). We argue that this particle then fused with the verb in Sadri to become a part of the new marker of the present tense. This marker was then reanalyzed in Sadri as *-l + =a*, with *-l* as part of the present-tense marker and =a denoting that the speaker considers the information of the respective clause to be especially relevant to the narrative. This last reanalysis was presumably motivated by speakers of Kherwarian languages using Sadri as a *lingua franca*, since a finite marker =a with a similar distribution is found in Kherwarian languages. While the function of this marker shows considerable speaker-specific variation in Sadri (and perhaps in Kherwarian as well), it is now a highly productive morpheme in Sadri, even in the speech of native speakers. This is all the more interesting since this category is only found in the present tense in Sadri, whereas it is found in all “finite” categories in Kherwarian, i.e., (simplifying somewhat) in declaratives and interrogatives in general, but not imperatives, rhetorical questions or non-indicative moods. It remains to be seen whether =a in Sadri will eventually spread to other tense/aspect categories, resulting in an even more similar distribution to Kherwarian.

Due to the discourse-pragmatic nature of this marker in Sadri and the small corpus on which our study is based, the present study can only be considered a first step towards understanding this marker in Sadri, and other factors which might play a role, such as prosody, still await investigation. There is also the issue of defining what EXACTLY “narrative relevance” is. Nevertheless, we believe that an analysis of =*a* in Sadri as a marker of narrative strategy best accounts for the available data and often even yields the same result as the homophonous finite marker in Kherwarian, thereby allowing us to posit a direct connection between the two. Thus, although the marker in Sadri is clearly of IA origin, there can be little doubt that its current distribution is due at least in part to that of the homophonous form in Kherwarian. In light of this discourse-pragmatic analysis of =*a* in Sadri, a review of the status of the homophonous marker in Kherwarian might also lead to interesting insights, as =*a* is generally portrayed as a fully grammaticized marker of finiteness in these languages. A fresh look at this marker in Kherwarian might even reveal that the “finite marker” =*a* in these languages is at least partially discourse-pragmatically oriented – perhaps due to Sadri influence.

Abbreviations

1, 2, 3	person
ABL	ablative
ACT	active
ADD	additive focus
ADJZ	adjectivizer
ANTIC	anticipatory
APPL	applicative
AUX	auxiliary
CAUS	causative
CLF	classifier
COMP	complementizer
COP	copula
CREL	corelative
DU	dual
EXT	extentional
FIN	finite
FOC	focus
FUT	future
GEN	genitive
HON	honorific status
IA	Indo-Aryan
IMP	imperative
INAN	inanimate

INCL	inclusive
IND	indicative
INF	infinitive
INGR	ingressive
IPFV	imperfective
LNK	linker
LOC	locative
MIA	Middle Indo-Aryan
MID	middle voice
NAR	narrative marker
NEG	negation
NIA	New Indo-Aryan
NMLZ	nominalizer
NPST	non-past
NSG	non-singular
NUM	number
OBJ	object
OBL	oblique case
OIA	Old Indo-Aryan
PASS	passive
PERS	person
PL	plural
POL	polite
POSS	possessive
PRS	present
PST	past
PTCP	participle
REFL	reflexive
SEQ	sequential converb
SG	singular
SBJV	subjunctive
SBJ	subject
TEL	telicizer
TOP	topic
TR	transitive
VOC	vocative
-W-	hiatus-breaking unit

References

- Abbi, Anvita. 1997. Languages in contact in Jharkhand. In Anvita Abbi (ed.), *Languages of tribal and indigenous peoples of India. The ethnic space*, 131–148. Delhi: Motilal Banarsidass.
- Aikhenvald, Alexandra Y. 2002. Typological parameters for the study of clitics, with special reference to Tariana. In R. M. W. Dixon & Alexandra Y. Aikhenvald (eds.), *Word. A cross-linguistic typology*, 42–78. Cambridge: Cambridge University Press.

- Anderson, Gregory D. S. 2007. *The Munda verb. Typological perspectives* (Trends in Linguistics, Studies and Monographs, 174). Berlin & New York: Mouton de Gruyter.
- Anderson, Gregory D. S. & Felix Rau. 2008. Gorum. In Gregory D. S. Anderson (ed.), *The Munda languages* (Routledge Language Family Series), 381–433. London & New York: Routledge.
- Anderson, Stephen R. 2005. *Aspects of the theory of clitics* (Oxford Studies in Theoretical Linguistics, 11). Oxford: Oxford University Press.
- Bloch, Jules. 1919. *La formation de la langue marathe* (Bibliothèque de l'École des Hautes Études, Sciences historiques et philologiques, 215). Paris: Champion.
- Chatterji, Suniti Kumar. 1926 [2002]. *The origin and development of the Bengali language. With a foreword by Sir George Abraham Grierson*. Delhi: Rupa.
- Deeney, J. S. J. 1975. *Ho grammar and vocabulary*. Chaibasa: Xavier Ho Publications, St. Xavier's High School.
- Ghosh, Arun. 1994 [2010]. *Santali. A look into Santal morphology*. New Delhi: Gyan Publishing House.
- Ghosh, Arun. 2008. Santali. In Gregory D. S. Anderson (ed.), *The Munda languages* (Routledge Language Family Series), 11–98. London & New York: Routledge.
- Hoffmann, Revd. J. S. J. 1905 & 1909 [2001]. *Mundari grammar and exercises. Parts I and II*. New Delhi: Gyan Publishing House.
- Jordan-Horstmann, Monika. 1969. *Sadani. A Bhojpuri dialect spoken in Chotanagpur* (Indologia Berolinensis, 1). Wiesbaden: Otto Harrassowitz.
- Lewis, M. Paul, Gary F. Simons & Charles D. Fennig (eds.). 2014. *Ethnologue: Languages of the world, seventeenth edition*. Dallas, TX: SIL International. Online version: <http://www.ethnologue.com> (accessed 19 February 2015).
- Masica, Colin P. 1991. *The Indo-Aryan languages* (Cambridge Language Surveys). Cambridge: Cambridge University Press.
- Neukom, Lukas. 2001. *Santali (Languages of the World/Materials, 323)*. München: Lincom Europa.
- Neukom, Lukas & Manideepa Patnaik. 2003. *A grammar of Oriya* (Arbeiten des Seminars für Allgemeine Sprachwissenschaft, 17). Zürich: Universität Zürich.
- Nowrangi, P.S. 1956. *A simple Sadāni grammar*. Ranchi: D. S. S. Book Depot.
- Osada, Toshiki. 1991. Linguistic convergence in the Chotanagpur area. In S. Bosu Mullick (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 Gregory D. S. Anderson (ed.), *The Munda languages* (Routledge Language Family Series), 99–164. London & New York: Routledge.
- Peterson, John. 2010. Language contact in Jharkhand. Linguistic convergence between Munda and Indo-Aryan in eastern-central India. *Himalayan Linguistics* 9(2). 56–86. <http://www.linguistics.ucsb.edu/HimalayanLinguistics/articles/2010/PDF/HLJ0902B.pdf> (accessed 19 February 2015).
- Peterson, John. to appear. Jharkhand as a “linguistic area” – Language contact between Indo-Aryan and Munda in eastern-central South Asia. In Raymond Hickey (ed.), *The Cambridge handbook of areal linguistics* (Cambridge Handbooks in Language and Linguistics). Cambridge: Cambridge University Press.
- Peterson, John & Savita Kiran. 2011. Sadani/Sadri. In L. I. Kulikov, T. I. Oranskaya & A. Yu. Rusakov (eds.), *Языки мира* (Languages of the world), vol. 16, 367–379. Moscow:

- Academia. English version: <http://www.southasiabibliography.de/uploads/Sadri.pdf> (accessed 19 February 2015).
- Pinnow, Heinz-Jürgen. 1966. A comparative study of the verb in the Munda languages. In Norman H. Zide (ed.), *Studies in comparative Austroasiatic linguistics* (Indo-Iranian Monographs, 5), 96–193. London, The Hague & Paris: Mouton.
- Pucilowski, Anna. 2013. *Topics in Ho morphophonology and morphosyntax*. University of Oregon dissertation. https://scholarsbank.uoregon.edu/xmlui/bitstream/handle/1794/13241/Pucilowski_oregon_0171A_10666.pdf?sequence=1 (accessed 19 February 2015).
- Ramaswami, N. 1992. *Bhumij grammar* (CILL Grammar Series, 18). Mysore: Central Institute of Indian Languages.
- Shukla, Shaligram. 1981. *Bhojpuri grammar*. Washington, D.C.: Georgetown University Press.
- Sinha, N. K. 1975. *Mundari grammar* (CILL Grammar Series, 2). Mysore: Central Institute of Indian Languages.
- Tiwari, Udai Narain. 1960. *The origin and development of Bhojpuri* (The Asiatic Society, Monograph Series, X). Calcutta: The Asiatic Society.

Note: The present study is largely based on data obtained during several visits to Jharkhand. I would like to thank the German Research Foundation (*Deutsche Forschungsgemeinschaft*) for generous grants which made three of these trips possible. I would also like to thank the Department of Tribal and Regional Languages at Ranchi University, and especially Dr. Ganesh Murmu, for their unwavering support over the years. Special thanks also go to the speakers of Sadri who freely gave of their time (and patience!) to explain the intricacies of their native language to me.