Netra P. Paudyal and John Peterson*

How one language became four: the impact of different contact-scenarios between “Sadani” and the tribal languages of Jharkhand

<https://doi.org/10.1515/jsall-2021-2028>

Published online May 5, 2021

Abstract: Four Indo-Aryan linguistic varieties are spoken in the state of Jharkhand in eastern central India, Sadri/Nagpuri, Khortha, Kurmali and Panchparganiya, which are considered by most linguists to be dialects of other, larger languages of the region, such as Bhojpuri, Magahi and Maithili, although their speakers consider them to be four distinct but closely related languages, collectively referred to as “Sadani”. In the present paper, we first make use of the program COG by the Summer Institute of Linguistics (SIL) to show that these four varieties do indeed form a distinct, compact genealogical group within the Magadhan language group of Indo-Aryan. We then go on to argue that the traditional classification of these languages as dialects of other languages appears to be based on morphosyntactic differences between these four languages and similarities with their larger neighbors such as Bhojpuri and Magahi, differences which have arisen due to the different contact situations in which they are found.

Keywords: Khortha; Kudmali; language contact; Sadani; Sadri

1 Introduction

While the first official language of the state of Jharkhand in eastern central India is Hindi, over 96% of the state population speaks a local tribal or regional language as their first (L1) or second language (L2) on a daily basis, and only 3.7% of the people speak Hindi as their first language (JTWRI 2013:4–5). “Tribal language” refers here to Austro-Asiatic and Dravidian languages spoken in the region, while

***Corresponding author: John Peterson**, Department of Linguistics, Institute for Scandinavian Studies, Frisian and General Linguistics (ISFAS), University of Kiel, Kiel, Germany, E-mail: jpeterson@isfas.uni-kiel.de

Netra P. Paudyal, Department of Linguistics, Institute for Scandinavian Studies, Frisian and General Linguistics (ISFAS), University of Kiel, Kiel, Germany, E-mail: n.paudyal@isfas.uni-kiel.de

“regional languages” are local Indo-Aryan languages restricted to this region which do not have any official status outside of Jharkhand. Sadri/Nagpuri and Khortha are the two largest regional languages spoken in Jharkhand, followed by Kurmali and Panchparganiya. Along with a few other languages, these four languages have been granted the status of second official state languages of Jharkhand. The four largest Munda languages spoken in Jharkhand are Santali, Mundari and Ho, all North Munda, and Kharia, a South Munda language. The largest Dravidian language spoken in Jharkhand is Kurux/Kurukh, which belongs to the North Dravidian branch together with the smaller Malto, also spoken in the state. Gondi (South-Central Dravidian) has also traditionally been spoken in some western parts of the state, although all members of this ethnic group in Jharkhand now appear to speak the Indo-Aryan language Sadri/Nagpuri as their L1.

“Sadani”, which is the main focus of the present study, is often found in the linguistic literature with one of two different meanings. The first, more restricted meaning refers to the linguistic variety now generally referred to in Jharkhand as either “Sadri” or “Nagpuri”, one of the four languages dealt with in the present study. For example, Nowrangi (1956) and Jordan-Horstmann (1969), two grammars of Sadri/Nagpuri, refer to this language in their works as *Sadani* and *Sadāni*, respectively. This use of the term *Sadani* appears to have been more prevalent in past decades than it is at present.

In the second, more general meaning, and that which reflects the situation in Jharkhand today, *Sadani* refers to the four closely related Indo-Aryan varieties Panchparganiya, Khortha, Kurmali and Sadri/Nagpuri collectively, which are spoken throughout Jharkhand and neighboring states. These four varieties are generally considered independent languages in the region. *Sadani* in this sense reflects the fact that these languages are traditionally spoken by the Sadan (*sādān*), the non-tribal, Indo-Aryan-speaking ethnic groups of Jharkhand (e.g. Nowrangi 1956: ii), who presumably first brought Indo-Aryan languages to this region.

In accordance with the use of these terms in Jharkhand today, we follow the second, more general use in this study, referring to *Sadani* as a cover term for all four closely related linguistic varieties traditionally spoken by the Sadan. Figure 1 shows the approximate relative positions of these four languages with respect to one another in Jharkhand.¹

All four of these languages belong to what has often been referred to as the “Bihari” branch of Indo-Aryan, a term going back at least to Grierson (1903) which includes the Sadani languages in our use of the term but also Bhojpuri, Maithili,

¹ Note that Figure 1 depicts the relative positions of these languages only in this region, the original homeland of all four varieties, although some of these, above all Sadri, are also spoken in other regions, e.g. in northeastern India. See Section 2.1 below.

Figure 1: The approximate relative positions of the four Sadani varieties of Jharkhand and neighboring states.³

and Magahi, among others. The term “Bihari” has fallen into disuse among most modern researchers, among other reasons² because it explicitly refers to the state of Bihar, now only half the size of what it was prior to 2000, when the southern half of that state became the newly formed state of Jharkhand, but also because these languages are also spoken in Chhattisgarh, eastern Uttar Pradesh, southern Nepal, western West Bengal and elsewhere. For this reason we prefer the term “Magadhan” to refer to these languages, as this refers to the larger region of eastern central India from which these languages originated, coinciding largely (although not entirely) with the territory of the former kingdom of Magadha.

In the usual classification of Indo-Aryan, at least assuming an “inner-outer” distinction,⁴ the Magadhan group is a member of the eastern outer group of Indo-Aryan, together with the Bengali–Assamese and Oriya branches. This is illustrated in Figure 2, adapted and simplified from the classification found in Eberhard et al.

² In addition, Grierson (1903) explicitly considers the members of his “Bihari” group to be dialects, a view which is now generally rejected.

³ The original map, which has been modified here somewhat, is copyrighted by “PlaneMad” [Arun Ganesh]/Wikipedia and can be downloaded at the following site (last accessed: 25 February, 2020): http://de.wikipedia.org/w/index.php?title=Datei:India_Jharkhand_locator_map.svg&filetimestamp=20081229064837. The changes that have been made are that the map shown here is only of Jharkhand and its direct surroundings, as well as the names of the languages given in the map.

⁴ The authors are entirely neutral with respect to the “Inner-Outer Hypothesis” and will not deal with this discussion further. Assuming its existence here is merely for ease of presentation.

(2019) for the Indo-Aryan group of languages, with a small number of languages indicated for reference.

Opinions among linguists differ considerably with respect to the status of the individual Sadani languages within Indo-Aryan. For example, whereas Eberhard et al. (2019) consider Kurmali and Panchparganiya to be separate languages, Grierson (1903: 146–147) considers them to be Eastern Magahi dialects. On the other hand, Khortha – which Grierson (1903) also considers to be an Eastern Magahi dialect – is listed as a dialect both of Maithili and Magahi by Eberhard et al. (2019). For Sadri/Nagpuri, there has been general consensus in studies produced by non-native speakers since at least Grierson (1903) that this is a Bhojpuri dialect (cf. e.g. Grierson 1903; Jordan-Horstmann 1969; Tiwari 1960), although Yadav (2012) suggests that Nepali Sadri might be a dialect of Maithili (2012: 2).

These classifications are at odds with the intuition of most native speakers, who view Sadri/Nagpuri, Khortha, Kurmali and Panchparganiya as languages in their own right and not dialects of neighboring languages. Instead, speakers almost unanimously view these four languages as *Sadani* languages, i.e., as closely related languages which descend from dialects of an earlier Sadani language, which no longer exists in that form. In the following section, we deal with this topic in detail and suggest, based on preliminary data, that the local tradition in fact very likely best represents the genealogical classification of these languages.

2 The status of Sadani within the Magadhan group

In order to determine to what extent the four Sadani languages form a distinct linguistic group or whether they are more closely related to neighboring Indo-Aryan languages of the region than to each other, we collected data together with native speakers for a slightly adapted version of the 100-item Swadesh list in field work for

Figure 2: The place of Magadhan within Indo-Aryan, adapted and simplified from. The classification in Eberhard et al. (2019).

nine distinct dialects of the Sadani languages as well as for various other eastern Indo-Aryan varieties, including two dialects each for Bhojpuri and Bengali (Bangla) and one variety each for Maithili, Magahi and Angika. Further data for the sake of comparison were included for other Indo-Aryan varieties from further afield, such as one non-standard variety of both Konkani (a husband and wife from Benaulim, South Goa) and Punjabi (one speaker from Delhi), data on three dialects of the Darai language of central Nepal from the first author's previous work, as well as standard Hindi and Nepali, totaling altogether 23 different varieties. Although this list of languages/dialects is admittedly primarily due to the availability of speakers that we were able to locate for languages not spoken in Jharkhand itself, at least one language from all three major divisions of Indo-Aryan, i.e., Intermediate and Outer Indo-Aryan as well as Western Hindi (cf. Figure 1 above), is included.

These raw data were transcribed in IPA and entered into tables. Loanwords from Persian/Arabic, English and Sanskrit were then removed. The identification of loanwords from other New Indo-Aryan languages proved to be somewhat more difficult, with the exception of the three Darai dialects, where Nepali loanwords were generally easily identifiable, although loanwords from Hindi in other languages which we were able to identify were also removed.

The loanwords which were removed from the raw data include highly common words of Persian origin (based on the data in McGregor 1997), such as Hindi (etc.) *rāstā* or *rāh* 'road', *lāl* 'red', *bes* (<*beś*) 'good', *dil* 'heart', etc., Sanskrit loans such as *onek* 'many' or *brifti* 'rain' in the two Bengali (Bangla) dialects, Ranchi Sadri *hirdai* 'heart', etc. Also excluded were entries which consisted of two or more words, such as Ranchi Sadri *kʰʌʎa hov-ek* 'to stand (up)' [standing become-INF], etc. Finally, as this sifting of the data occasionally resulted in some headwords having entries only for a few languages, where most of the original entries were loanwords (e.g. *dil* for 'heart' or *lāl* for 'red' in numerous languages), all headwords with fewer than 10 cognate forms were removed from the data. This reduced the number of headwords from the original 103 to 94.

The data thus collected and filtered were then analyzed with the software COG (COG 2019) by the Summer Institute of Linguistics (SIL),⁵ the results of which are presented in Figures 3–6. These illustrate the analysis of the relations between the various linguistic varieties in both a UPGMA and Neighbor-Joining⁶ analysis with respect to both lexical and phonetic similarity.

⁵ <https://software.sil.org/cog/> (last accessed: 26 February, 2020).

⁶ For reasons of space, we will not deal here further with the different assumptions made in the UPGMA vs. Neighbor-Joining analyses. Further discussion of these and similar approaches can be found in Nichols and Warnow (2008). For our purposes it will be enough to simply consider these to be two competing approaches to analyzing the data.

Figure 3: The genealogical relationship of various “Magadhan” languages and dialects with respect to selected Indo-Aryan varieties – Lexical similarity, UPGMA.

In Figure 3, a UPGMA analysis of lexical similarity, we find three main branches within a clearly defined, compact Sadani group. In the first (the upper one of the three), Kurmali⁷ diverges from the remaining Sadani languages. This is expected since the etymology of many common lexemes in this language is unclear, and they likely represent loanwords or substrate influence from a non-Indo-Aryan/non-Dravidian/non-Munda language which is no longer spoken (see Section 3.3). Consider in this respect e.g. *nuṛ-* ‘eat’ or *tʰanau-* ‘see’ vs. the more common Magadhan (and Indo-Aryan generally) forms *kʰa-* and *dekʰ-*, respectively, to cite just two such examples. Following the divergence of Kurmali from the rest of Sadani, we then find a two-way split between all Sadri/Nagpuri dialects and Panchparganiya on the one hand vs. the Khortha dialects on the other, which corresponds well with these authors’ own impression that Panchparganiya and Sadri are so closely related that they are perhaps best analyzed as two highly similar dialect groups, distinct from Khortha. Thus, the Sadani languages form a coherent group in Figure 3 within an equally tightly knit Magadhan group, and all Sadani languages are more closely related to one another than to any language(s) outside of this group. Outside of Magadhan, the closest relatives are the two Bangla dialects, as is to be expected. Hindi, Punjabi and Nepali also form a separate group. Konkani is, as expected, quite

⁷ Also written as <Kuṛmali>, <Kuṛmali>, <Kuḍmali> or <Kudmali>.

Figure 4: The genealogical relationship of various “Magadhan” languages and dialects with respect to selected Indo-Aryan varieties – Lexical similarity, Neighbor-joining.

Figure 5: The genealogical relationship of various “Magadhan” languages and dialects with respect to selected Indo-Aryan varieties – Phonetic similarity, UPGMA.

Figure 6: The genealogical relationship of various “Magadhan” languages and dialects with respect to selected Indo-Aryan varieties – Phonetic similarity, Neighbor-Joining.

distant from the other languages in this tree, and the three Darai varieties occupy a position between Konkani and the rest of Indo-Aryan.

In contrast, in the Neighbor-Joining analysis of lexical similarity in Figure 4, the respective positions of Maithili, Angika, Magahi and the two Bhojpuri varieties are somewhat different from their positions in Figure. However these languages still form a clear Magadhan group which also encompasses the Sadani languages, which here as in Figure 3 form a distinct group. Note however the position of Panchparganiya within the Khortha group in Figure 4, which is highly counter-intuitive considering that it is viewed by many (including these authors) as so close to Sadri as to almost form a dialect of this language. Kurmali is also placed within the Khortha group here, again a rather counter-intuitive position considering its divergent status within Sadani. Hence, on both of these accounts, the UPGMA tree produces more expected results than the Neighbor-Joining analysis, although in both analyses Sadani forms a distinct subgroup within a coherent and distinctive Magadhan group. Outside of Magadhan, the only major difference to Figure 3 is that Nepali here no longer clusters with Hindi and Punjabi but rather with Darai, again a rather unexpected position.

Figure 5, which shows a UPGMA analysis of phonetic similarity, is virtually identical with Figure 3, the UPGMA analysis of lexical similarity, with only slight differences with respect to the internal structure of the Sadri dialects, and with

respect to Nepali. With respect to the status of Sadani, we note once again that a clear-cut, tightly knit Sadani group within – but also distinct from – a larger, equally distinct Magadhan group is again confirmed by the analysis, and all Sadani languages are genealogically more closely related to one another than to any other languages.

Finally, Figure 6 presents the Neighbor-Joining tree for phonetic similarity. Here again we find a clearly distinguishable Sadani group within an equally distinct Maghadhan group, again with a somewhat counter-intuitive position within the Sadani group for both Kurmali and Panchparganiya, as in Figure 4. However, as in Figures 3–5, Sadani in Figure 6 forms a clearly distinct group within Magadhan, and all Sadani languages are genealogically more closely related to one another than to any Magadhan (or other) language.

All four analyses in COG thus confirm the unity of Sadani as a distinct branch within Magadhan, despite all the finer differences shown by the different analyses, such as the position of Nepali, or the differences in the internal hierarchical structure of Magadhan and Sadani.

A closer look at the cognate and non-cognate forms proposed by the program COG reveals a number of problems, however. To begin with, all first persons singular were viewed as cognate with the exception of Konkani *hāu*, which represents an older nominative form no longer found in the other languages included in this sample, which go back to the earlier ergative. As such, its status as “non-cognate” in COG is to be expected. Problematic is the fact that forms such as Magahi or Maithili *ham* or *hām*, respectively, which derive from the historically plural first person forms, were also viewed as cognates with the remaining first persons singular, which derive from the historically singular forms, such as Jaldega Sadri *mṣē* or Hindi and Punjabi *mē*. Also unexpectedly, the two Bengali forms for the first person plural ‘we’, both *amra*, were analyzed differently: Kanchrapara Bangla *amra* was analyzed as cognate (but forming its own cognate set) with forms in other languages such as Nepali *hami* or Jaldega Sadri *hame*, but Southern Bangla *amra* was not considered cognate with the identical form in Kanchrapara Bangla.

To cite a few other examples, while forms for ‘bone’ such as *haṛ* and *haḍḍi* were correctly analyzed as cognates, Benaulim Konkani *haḍ* was not viewed as cognate. Similar problems were encountered elsewhere, e.g. of the various forms for ‘bark (of a tree)’, several forms such as *ts^hala(k)*, *ʃ^hal*, *sal* and *ts^hilka* were considered cognate while the (Delhi) Punjabi form *ʃ^hal* oddly was not. Similarly, forms such as *ninda* ‘sleep’, found in various varieties, were generally recognized as cognates, cf. Garhwa Sadri, Ranchi Sadri and Jaldega Sadri *ninda-ek* (where *-ek* is the infinitival marker), which were analyzed as cognates in one cognate set, and Panchparganiya *nida-e*, with the infinitival marker *-e*, in a different cognate set. However, the almost identical Garhwa Sadri alternate form *ninda-e* was not considered cognate with these forms.

It is thus clear that automated tools such as COG at present cannot fully substitute for a careful manual analysis of the data within the traditional comparative method of historical linguistics by human researchers, e.g. identifying and removing loanwords and identifying (non-)cognates, etc., and COG does produce a number of false (non-)cognates. Nevertheless, as this study also shows, such a tool can be enormously useful in obtaining quick, preliminary results which can then guide further research, and although a number of false analyses did occur, the number of correct analyses was apparently large enough to balance these out, as at least the two UPGMA analyses were largely in line with these authors' intuitions and with local opinion in Jharkhand, while the Neighbor-Joining analyses were somewhat less reliable. Automatic analyses such as those offered by COG thus present us with a further valuable tool whose potential benefit should not be under-estimated, especially when combined with traditional methods of historical linguistics.

In summary, in all of these different analyses, each with its own advantages and disadvantages, the Sadani group consisting of the languages Sadri/Nagpuri, Khortha, Kurmali and Panchparganiya is clearly recognizable as a distinct group contained within an equally distinct Magadhan group. This lends strong support to the locally held view that Sadri/Nagpuri, Panchparganiya, Khortha and Kurmali are four closely related languages, deriving from a common ancestor and distinct from other Magadhan languages.

The fact that these languages form their own distinct group and are more closely related to each other within the Sadani group than to any language or languages outside of this group thus strongly suggests that Sadri/Nagpuri cannot be viewed as a dialect of Bhojpuri and Panchparganiya, and Khortha and Kurmali cannot be viewed as dialects of Magahi, Maithili or any other language. Rather, as we argue in the following pages, this traditional linguistic classification of the Sadani languages as dialects of other Magadhan languages appears to be motivated by morphosyntactic aspects, such as the presence or absence of ergativity, or mono- vs. polypersonal verb marking, to which we now turn.

3 The Sadani languages, their speakers and their distinguishing features

In addition to the lexical and phonetic similarities between the Sadani languages discussed in Section 2, there are also a number of common structural features in all

four Sadani languages, many of which are shared with most other eastern Indo-Aryan languages. To name just a few:

- Nouns in all four Sadani languages have one invariable base, i.e., there is nothing in these languages comparable to the oblique stem found in western Indo-Aryan languages which is obligatory before postpositions, while the direct stem (“direct/nominative case”) is found elsewhere.
- Almost all grammatical marking within the NP such as number and case is enclitic, not suffixal, as it is in western Indo-Aryan languages such as Hindi, Marathi, or Konkani.
- All four Sadani languages have numeral classifiers, most of which can also be used post-nominally to mark the NP as definite or referential.
- All four Sadani languages also have a vigesimal system (e.g. *kori* or *kath* ‘score; 20’), likely from Austro-Asiatic, often in addition to the inherited Indo-Aryan form *bis* ‘20’.

Nevertheless, the four members of the Sadani group also differ from one another somewhat with respect to the lexicon and morphosyntax. For example, some languages have morphological ergativity while others do not; some show an alienable/inalienable distinction with attributive possession while others do not; some have polypersonal marking on the verb (i.e., agreement with more than one argument/speech-act participant) while others do not, etc. As we argue in Sections 3.1–3.4, such differences most likely arose through the different contact scenarios that each of these four languages finds itself in.

In the following sections we present a brief overview of some of the distinctive features of each individual language and its respective contact situation, in an attempt to account for the differences between the four languages. For ease of reference, Figure 7 provides an overview of the approximate relative positions of the various Munda languages of the region, from Anderson (2007: 7), many of which are also contact languages of the Sadani group. Kurukh/Kurux, the main Dravidian language of the region (not shown in Figure 7), is predominantly spoken throughout the northern half of western Jharkhand, corresponding roughly with the northern half of the Sadri-speaking area, but also in contact with all three other Sadani languages to a lesser extent (see again Figure 1).

3.1 Sadri/Nagpuri

Sadri/Nagpuri is spoken in western and central Jharkhand and in neighboring regions of Chhattisgarh and Odisha. Other varieties of Sadri are spoken in Assam

Figure 7: The approximate relative positions of the Munda languages (from Anderson 2007: 7, reprinted with kind permission by Mouton de Gruyter).

and West Bengal, the Andaman and Nicobar Islands, in Bangladesh, southern central Nepal, and in Bhutan. *Sadani* and *Nagpuri* refer to two somewhat different, register-based forms of this language. Generally speaking, *Sadri* refers to the spoken, non-literary form of this language, especially the language spoken by tribals in the countryside, while *Nagpuri* refers to the polished, literary language, especially as used by Hindus and in cities.

Sadri/Nagpuri is the major *lingua franca* of western and central Jharkhand for many tribal groups, both Munda and Dravidian. These include speakers of Kharia (South Munda), Mundari, Bhumij, Turi, Asur, Birhor, Korwa, Koduku, Bijori, etc. (North Munda), Kurukh (North Dravidian) and Gondi (Central Dravidian), and large numbers of members of these ethnic groups now speak Sadri as their first language and no longer speak their traditional languages (Peterson and Baraik, In press).

According to the Ethnologue (Eberhard et al. 2019), varieties of Sadri are spoken in India by 12,130,000 people, of whom 5,130,000 speak it as a first language (L1 speakers) and 7,000,000 as a second language (L2 speakers). The total given there for all countries is 12,131,225, with 5,131,180 L1 speakers and a

further 7,000,045 L2 speakers. Second-language speakers thus account for a clear majority of all Sadri/Nagpuri speakers, making up ca. 58% of the total both in India and in all countries. With that, Sadri/Nagpuri is the second largest Sadani language with respect to the number of L1 speakers, after Khortha (see Section 3.2), but the largest with respect to the total number of speakers, i.e., L1 and L2 speakers combined.

Due to its status as a *lingua franca* for many non-Indo-Aryan-speaking ethnic groups, Sadri/Nagpuri has been greatly affected by the Munda and Dravidian languages spoken in Jharkhand and neighboring regions. We now turn to some of its characteristic traits.

Sadri/Nagpuri undoubtedly has a large number of loanwords from different languages in the area, although no systematic study of these has yet been undertaken. However it does not appear to have a large number of loanwords from one major source, unlike e.g. Khortha and Kurmali, as discussed in Sections 3.2 and 3.3. Instead, it itself is a major source for countless loanwords for its contact languages, which constitute a considerable portion of those languages' lexica, including core areas such as numerals, kinship terms, body parts, etc. (Abbi 1997).

Plural marking in the NP in Sadri/Nagpuri, unlike in other Sadani languages, is always via the enclitic =*man*, both with nouns and pronouns. As there is only one plural marker in Sadri/Nagpuri compared with various markers in the other Sadani languages, this can be considered a kind of "simplification" of the grammar, due to Sadri/Nagpuri's status as a *lingua franca*.

Sadri/Nagpuri has no morphological ergativity. With the exception of a number of experiential predicates,⁸ it is a "nominative/accusative language" in which S and A consistently appear in the unmarked or "direct" case and P (or O) is either unmarked, when its reference is non-specific and/or non-human, or marked by the oblique case (=ke) when it is definite and/or human. If we assume that Sadani originally had morphological ergativity when it was first brought to what is now Jharkhand – a reasonable assumption, since it is still found in the other three Sadani languages (see Sections 3.2–3.4) – then its loss in Sadri/Nagpuri could be accounted for through the status of this language as a *lingua franca* by speakers of other languages, none of which has ergativity, whether North Munda (Mundari, Ho, Turi, etc.), South Munda (Kharia), North Dravidian (Kurux) or Central Dravidian (Gondi).

⁸ With experiential predicates, the experiencer can be unmarked (= "nominative"), or it can appear in the genitive or oblique case (the latter covering roughly the same functions as the dative and accusative in many languages).

Only one argument is marked on the predicate in Sadri/Nagpuri – the S of the intransitive predicate or the A of the transitive predicate, but never P (O). This is different for Khortha and Kurmalī, both of which have polypersonal marking to some extent, as do Magahi, Maithili and other languages in the region as well. In Peterson (2017) it is suggested that polypersonal marking likely arose in situations of intense contact between proto-Munda and Middle Indo-Aryan in this general area, as it is not found in the Dravidian languages of the region such as Malto or Kurux, cf. Kobayashi (2012) and Kobayashi and Tirkey (2017), respectively.

As Sadri/Nagpuri is in close contact with North Munda, where polypersonal marking is the norm, it may at first seem somewhat mystifying that it is not found in Sadri/Nagpuri as well. However Sadri/Nagpuri is also used as a *lingua franca* by speakers of Kharia (South Munda), Kurux (North Dravidian) and Gondi (Central Dravidian), none of which have polypersonal marking. As will become evident in the following sections, polypersonal marking in Sadani languages is likely an innovation, being found only in those two languages – Khortha and Kurmalī – which are in close contact with one North Munda major contact language which itself possesses polypersonal marking. In this sense Sadri/Nagpuri was conservative, as it did not develop this feature.

One characteristic feature of Sadri/Nagpuri which is not found in other Sadani languages is that it distinguishes between alienable and inalienable attributive possession. This distinction is made only with third person possessors, not with first or second person possessors. Cf. (1): here the enclitic genitive marker =*ak* denotes that the possessive relationship between the owners of the field and the field itself (i.e., ‘their field = the field of these’) is an instance of alienable possession.

- (1) *e=he=man=ak doin me kʌr-ek kʌh-b-e!*⁹
 this=FOC=PL=GEN field in do-INF say-FUT.IMP-2SG
 ‘You tell him to do [that] in their field!’

The morph =*har*/=*hʌr* on the other hand marks an inalienable possessive relationship, as (2) shows. It is quite commonly found (but never obligatory) with kinship terms (2), body parts (3), and a few other lexemes, such as ‘car’ (4).

⁹ Examples from the authors’ own corpora are given without any source; only the sources of examples taken from other published works are given in the text. Furthermore, glosses and translations in examples from other authors have been added in the present study when none were present in those works, or have been reanalyzed when deemed appropriate. These minor changes are not indicated in the examples.

- (2) *ghar ghur-l-ak to janana=har puch-l-ak, ...*
 house return-PST-3SG TOP wife=3POSS ask-PST-3SG
 'He returned home and his wife asked ...'
 (Nowrangi 1956: 161)
- (3) *ghari to coila=har=ke khapke dhar-l-õ.*
 time TOP tail.of.scorpion=3POSS=OBL quickly grab-PST-1SG
 'Then I quickly grabbed the scorpion's tail.'
- (4) *aur gari=har*
 and car=3POSS
 'and his car'

An alienable/inalienable distinction in attributive possession is otherwise not found in Indo-Aryan languages of the region, but is found in both North and South Munda languages for whose speakers Sadri/Nagpuri is the traditional *lingua franca*. We therefore assume that this trait has entered Sadri/Nagpuri through the use of this language as a *lingua franca* by L2 speakers of these North and South Munda languages.

Another characteristic feature of Sadri/Nagpuri is the narrative marker =*a*, which is found in the present tense and denotes that the information contained in the clause whose predicate is marked by =*a* contains information which the speaker portrays as especially relevant. This marker derives from the second half of the marker *=*la*, most likely originally a focus marker, which is found in many Indo-Aryan languages forming part of the present or future paradigms, at least for some persons. In Sadri/Nagpuri, the /l/ was reanalyzed as forming part of the present tense, while =*a* was reanalyzed as a narrative marker.¹⁰ Table 1, based on Nowrangi (1956: 90–91), illustrates this marker and its relation to the present tense. /n/ instead of /l/ in the first person singular and third person plural is due to the presence of a nasalized vowel or /n/ preceding earlier *=*la*, and in the second person singular, /l/ is also omitted when the narrative is not marked, as /sl/ is not a possible consonant cluster in word-final position.

Peterson (2015) argues that the use of =*a* in Sadri as a narrative marker ultimately derives from North Munda languages such as Mundari and Ho, with which Sadri is in close contact. In these and other North Munda languages, =*a* marks a predicate form as finite, i.e., as the predicate of the main clause (with some exceptions), but does not mark the predicate of subordinate clauses or rhetorical

¹⁰ See already Jordan-Horstmann (1969: 79) on the reanalysis of /l/ as part of the present tense marker. For further information on the etymology of this marker, cf. Bloch (1919: 241, §242); Chatterji (2002 [1926]: 997, §728); and Tiwari (1960: 176, §588).

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

Person	Singular	Plural
1	<i>khaōn(a)</i>	<i>khail(a)</i>
2	<i>khais(la)</i>	<i>khawal(a)</i>
3	<i>khael(a)</i>	<i>khaen(a)</i>

questions. Although this function differs from that in Sadri, as Peterson (2015: 196) notes, the distribution of *=a* in actual texts closely resembles that found in North Munda. (5) illustrates the “finite” marker *=a* in Mundari, while (6) shows the use of the narrative marker *=a* in Sadri. Note that the scope of the final *=a* in (6) extends to both of the two last verbs in the example, showing the enclitic nature of this marker.

(5) Mundari

*ranci do=jɪ **lel-aka-d=a***
 Ranchi TOP=1SG see-CONT-TR=FIN
 ‘I have seen Ranchi; I have been to Ranchi.’
 (adapted from Osada 2008: 127)

(6) Sadri/Nagpuri

*bec=le se paisa **mil-el=a** ʌur u=kʌr se hamʌr*
 sell=ANTIC then money be.found-PRS.3SG=NAR and that=GEN ABL 1PL.GEN
ʌdmi=mʌn sag sʌbji cawʌl dail ʌur
 man=PL spinach vegetable unhusked.rice lentil.soup and
le=ke
 take=SEQ
an-en=a** ʌur **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 **eat [and] drink.**’

In summary, we note a number of features in Sadri/Nagpuri which appear to have been motivated by its use as an L2 by speakers of many different languages. In the following sections we take a closer look at the three remaining Sadani languages and compare these with Sadri/Nagpuri as discussed in this section, but also with their respective linguistic neighbors.

3.2 Khortha

Khortha is the largest Sadani language in terms of the number of L1 speakers. According to the census report of 2011 (GOI 2011), Khortha has 8,038,735 L1 speakers. It is also used as a *lingua franca* in the central part of the Jharkhand. The major tribes who speak Khortha as an L2 include the Mundari of north Chotanagpur and the Santali of the Santhal Parganas and north Chotanagpur, Bedi(y)a, Lohra, Kharwar, as well as a limited number of Oraon (or Kurux), living in Bokaro. L1 speakers of Santali constitute the largest number of L2 Khortha speakers (Paudyal in progress).

Khortha speakers are scattered throughout all of North Chotanagpur, the Santhal Parganas and the Palamu division. The districts of Bokaro, Ramgarh and Hazaribagh are considered to be the core Khortha-speaking areas of Jharkhand. These form the central part of the state and the majority of Khortha speakers live in these areas.

As Khortha is the *lingua franca* of a large number of Santali speakers (North Munda), there are many common lexical borrowings from Santali in Khortha. (7) presents a few of these.

- (7)
<i>gidar</i>	‘child’
<i>kaṛa</i>	‘buffalo (male)’
<i>koṛi</i>	‘digger’
<i>casa</i>	‘farmer’

Khortha also makes extensive use of Santali borrowings in compounds. Here the first member of the compound is Santali while the second member is from an Indo-Aryan language, usually Khortha but occasionally also from Magahi. The examples in (8) illustrate this.

- (8)
<i>gidar-butru</i>	[child-child]	‘children’ (Santali-Magahi)
<i>koṛi-kudair</i>	[spade-spade]	‘spade’ (Santali-Khortha)
<i>gaṛa-dōṛa</i>	[ditch-ditch]	‘ditch’ (Santali-Khortha)
<i>casa-bari</i>	[cultivation-farm]	‘farm’ (Santali-Khortha)

Unlike Sadri/Nagpuri, Khortha has both a periphrastic and morphological number-marking system. For example, the enclitics =*gula* and =*guli* are productively used to mark the plurality of nouns periphrastically, but a number of nouns also inflect with various suffixes to indicate plurality. The most frequently occurring plural suffix markers are: -*ain*, -*wain*, -*lain*, and -(i)*an*. Unlike the enclitics =*gula/guli*, with these suffixes an/a/ or /ai/ in the singular stem is realized as /ʌ/ in the plural stem, as in e.g. *ghar* ‘house’ → *ghar-wain* ‘houses’ and *gaich* ‘tree’ → *gach-wain* ‘trees’. Examples

of both periphrastic and morphological number marking are illustrated in (9) and (10). Speakers generally prefer the enclitic markers over the suffixes.

- (9) *gidar* ‘child’ *gidar-lin* ‘children’¹¹
tūŋri ‘hill’ *tūŋri-lin* ‘hills’
ghar ‘house’ *ghar-wain* ‘houses’
gaich ‘tree’ *galch-wain* ‘trees’

- (10) *casa* ‘farmer’ *casa=gula* ‘farmers’
patai ‘leaf’ *patai=gula* ‘leaves’

Khortha has a moderate number of both native and non-native classifiers to mark definiteness. In addition to the more common forms =*ta*, =*ti*, =*tho*, =*go/gor*, =*muɽ*, and =*har* found in other Sadani languages, we also find the highly productive classifier =*gāɽa* [CLF] to denote units of four, borrowed from Santali *gandā*. A cognate form is also found in Mundari, *gand*, with the same function.¹² Consider (11)–(12) from the Khortha corpus and (13)–(14) for similar examples in Santali.

- (11) *murgi=ti* *du=gāɽa* *ɽim* *de-l=hae*.
 chicken=CLF.F two=CLF egg give-PST=AUX.PRS.3SG
 ‘The hen has laid eight eggs.’

- (12) *okar* *ek=gāɽa* *chluwa* *hae*.
 3SG.GEN one=CLF child COP.PRS.3SG
 ‘He has four children.’
 Santali

- (13) *pon=gandā* *aphor* *menak’=a*.
 four=CLF rice.seedlings COP=FIN
 ‘There are sixteen bunches of rice seedlings.’
 (Campbell 1899:177)

- (14) *bar=gandā* *sim* *hopon* *menak’=ko=ta=le=a*.
 two=CLF hen child COP=3PL=POSS=1PL=FIN
 ‘We have eight chicks (at our home).’

¹¹ *gidar* ‘child’ is a loan from Santali, cf. Santali *gidra* (see Basic word list of Santali; Neukom 2001: 224). It is not found in Sadri/Nagpuri and Kurmali. Unlike the Indo-Aryan term *chluwa* ‘child’, *gidar* is not used with non-humans, for example **chagair gidar* [goat child] ‘baby goat’, while *chagair chluwa* ‘baby goat’ or *kukur chluwa* ‘puppy’ are grammatical.

¹² We thank the audience at the International Workshop on the Documentation of Endangered Languages and Cultures, with special reference to Jharkhand, at the Dr. Shyama Prasad Mukherjee University on 11th and 12th April, 2019, for pointing out that *gāɽa* derives from Munda.

Unlike Sadri/Nagpuri, discussed above, most dialects of Khortha have an ergative case, =*ẽ*, which is limited to transitive subjects expressed by nouns or the first person singular and which is only found in the perfective aspect. Cf. (15)–(18).

- (15) *Budhna kail apɒn bhai=ke (thẽnga=ẽ) dheɽaw-t-ɒi.*
 PN tomorrow his brother=OBL (stick=INS) beat-FUT-3SG
 ‘Budhna will beat his brother tomorrow (with a stick).’
- (16) *Budhna=ẽ aij bihane sãp=ke (thẽnga=ẽ) dheɽaw-l-ɒi.*
 PN=ERG today morning snake=OBL stick=INS beat-PST-3SG
 ‘Budhna killed the snake this morning (with a stick).’
- (17) *mɒŋru=ẽ chagɒir=tɨ=ke budhwa=ke bech-l-ɒi.*
 PN=ERG goat=CLF=OBL PN=OBL sell-PST-3SG
 ‘Mangru sold the goat to Budhwa.’
- (18) *ham=ẽ piji tin bɒchar-ẽ pura kar-l-i=o.*
 1SG=ERG P.G. three year-LOC finishing do-PST-1=ADDR
 ‘I finished (my) post graduate in three years.’

However, not all dialects of Khortha have morphological ergativity, such as Parndiya, spoken in Hazaribagh, Chatra and Giridih districts. For example, in these dialects (18) would be as in (19).

- (19) *ham piji tin bɒchar-ẽ pura kar-l-õ.*
 1SG P.G. three year-LOC finishing do-PST-1SG
 ‘I finished (my) post graduate in three years.’

In Khortha, as in many other Indo-Aryan languages as well, the ergative marker is homophonous with the instrumental, as (15)–(16) above and (20) below show.

- (20) *phulkobi=k bihin ɾɒud=ẽ mɒ<i>r ge-l-ɒi.*
 cauliflower=GEN plant sun=INS die<LNK> go-PST-3SG
 ‘The cauliflower plant died from (= “through”) the sun.’

Also unlike Sadri/Nagpuri, Khortha has polypersonal marking when the subject is a first person and the object is a third person, i.e., $1 \rightarrow 3$. Consider (21)–(22), where the verb agrees with both A and P.

- (21) *ham okra dekh-l-i-ɒi.*
 1SG 3SG.OBL see-PST-1-3SG
 ‘I saw him.’

- (22) *kail=e bap=ke ḍagḍar=ke dekha-i de-l-i-ṇi.*
 yesterday=FOC father=OBL doctor=OBL show-LNK V2-PST-1-3SG
 'It was yesterday that I showed father to the doctor.'

Polypersonal marking in Khortha is likely part of a larger areal feature. It is found for example both in Magahi (Indo-Aryan) to the north as well as in Santali, its largest contact language. (23)–(24) illustrate polypersonal marking in Santali. One major difference between Santali and Khortha is that in Santali, in addition to the subject, both direct and indirect objects are marked on the verb, as shown in (23)–(24), whereas in Khortha object marking is restricted to direct objects, not indirect objects. The subject enclitic often attaches to the last unit preceding the predicate in Santali if the sentence consists of more than one word, as in (23), although not obligatorily, as (24) shows.

- (23) *teheṇ=dɔ=ṇ ṇam-akat'=ben=a.*
 today=TOP=1SG.SUBJ find-PRF:ACT=2DU.OBJ=IND
 'Today I have found you two.'
 (adapted from Neukom 2001:116)

- (24) *thoṛa daka juṭuc'-a=ṇ=me.*
 little rice add-APPL=1SG.OBJ=2SG.SUBJ
 'Give me a little more rice.'
 (adapted from Neukom 2001:50)

Unlike Kurmali, which also has polypersonal marking (Sections 3.3 and 3.4), in Khortha this polypersonal marking also extends to include addressee marking to express the relation between speaker and addressee. Addressee marking is a regular feature attested both in day-to-day speech and in written texts in Khortha. What differentiates addressee marking from the polyargument marking generally associated with polypersonal marking is that in addressee marking the relation between the speaker and addressee is marked on the predicate without the addressee necessarily being a participant in the event. We assume here that polypersonal marking in Khortha was originally restricted to subjects and direct objects and that it later spread to include addressees as well.

In the following example, one of our consultants is talking about his experience with his father. The addressee is marked here by *=o* and would not be used e.g. if this speaker were writing this in his diary, i.e., in a situation with no clear addressee.

- (25) *suru jhika bon ge-l=ha-l-i=o.*
 first turn forest GO-PST=AUX-PST-1SG=ADDR
se pura tha<i>k=ke leslamuṛ bhae=ge-l=ha-l-i=o.
 so completely tired<LNK>=CVB like.a.log COP=V2-PST=COP-PST-
 1=ADDR

‘I had gone to the jungle for the first time, so I had become very tired like a log.’

Addressee marking is also attested in the neighboring Magadhan language Magahi, but not in any other Sadani language, to our knowledge. We assume that it is for reasons such as this that Khortha has generally been considered a Magahi dialect by linguists.

3.3 Kurmali

Kurmali (also written Kudmali, Kuṛmali, Kuṛmali or Kuḍmali) is the traditional language of the Kurmi people. The majority of Kurmali speakers live in Jharkhand but there are also Kurmali-speaking communities in the neighboring states of Odisha, West Bengal and Assam. According to the 2011 census report (GOI 2011:6), Kurmali is spoken altogether by 311,175 people. However, almost all native writers and speakers we spoke to believe that the actual number of Kurmali speakers is far higher than the number cited in the census. Unlike Sadri/Nagpuri and Khortha, Kurmali is not commonly used as a *lingua franca*, although there are a few groups of Mundari and Santali speakers who do use it as for inter-ethnic communication.

Although Kurmali is certainly an Indo-Aryan language, the ethnic background of the Kurmi people is currently a matter of intense debate. Along with several other tribes living in the earlier provinces of Bihar and Orissa during the rule of the East India Company (or “British Raj”), the Kurmi people were classified and listed as tribals. However, just a few years after independence in 1950 they were withdrawn from the list of tribes and placed in the category of “non-tribal people”. Since then there have been regular demonstrations by the Kurmi people demanding their reclassification as a tribal group.

Whether the Kurmis should be considered ethnic Sadans, like the Khortha and most of the Sadri/Nagpuri speakers, or as tribes is beyond the scope of this paper. However, there are considerable cultural differences between the Kurmis and the (other) Sadan groups, such as the Kurmis’ complex clan system, with 81 different clans, each with a separate totem, which is not typical for ethnic Sadan but is more typical of the tribal groups of the region. Most of the clan names end in *-ar*, and the respective totem is generally a plant, an animal, or a type of fruit. Also similar to the

tribal groups of the region, persons belonging to a particular totem do not consume that totem, e.g., the *kēkuār* or ‘crab clan’ do not kill or eat crabs, and a member of the *kāhrṛār* or ‘pumpkin clan’ does not plant or eat pumpkins.

Kurmali also has many common lexical items that do not appear to be of Indo-Aryan origin but are also apparently not of Munda or Dravidian origin, including common lexemes such as *nuṭ*- ‘to eat’, *thanau*- ‘to see’. (26) presents a number of these from our Kurmali corpus.

- (26) *ankhai* ‘very’
 bṛṇa ‘inspect’
 brɔncɔ ‘instead of’
 ḍula ‘cut (esp. trees)’
 gijɔɾek ‘to laugh’
 gucek ‘open’
 jalaĩ ‘for’
 nuṭ- ‘to eat’
 sakar ‘pile of dust’
 sɔɾgek ‘sleep’
 tuɾek ‘write’
 thomkek ‘to wait’
 thanau- ‘to see’
 usas ‘easy’

There are also a number of grammatical markers and categories in Kurmali which are not found in other Sadani languages. For example, the plural marker for nouns is =*gila*/=*gili* for male and female, respectively, which is not unusual for Indo-Aryan languages of this region. However, unlike in other Sadani languages there is also a separate enclitic marker for the associative plural, =*nikha*, of unknown origin. Kurmali (like Panchparganiya, see Section 3.4) also marks the locative case with the enclitic =*mahan*, which does not appear to be of Indo-Aryan origin.

It thus appears that the Kurmi people once spoke a distinct language, neither Munda nor Dravidian but also not Indo-Aryan, and at some point switched to the regional Indo-Aryan *lingua franca* of that time, leaving a distinct substrate in their new language. Further research into this topic is likely to produce interesting results.

Kurmali also has a number of lexical borrowings from Santali, although not as many as Khortha. These include forms such as *bejaĩ* ‘very’, likely also *iṛkek* ‘run away’ (<Santali *iḍi* ‘take away’?) and perhaps also *nistai* ‘exactly, truly’ (<Santali *niṭsahi* ‘exactly/truly’). There is also a productive derivational prefix in Kurmali, *ḍɔr*- ‘half, incomplete, insufficient’, which likely derives from the Santali adjective *dorasa* ‘slightly’. A few examples are given in (27).

- (27) *pēdɔla* ‘lie’ *dɔrpēdɔla* ‘the person who mostly lies’
pɔdha ‘study’ *dɔrpɔdha* ‘the one who didn’t complete his/her studies’
haga ‘defecate’ *dɔrhaga* ‘without fully defecating’
pura ‘complete’ *dɔrpura* ‘incomplete’

Further possible influence from Santali (or perhaps the unknown language presumed to have once been spoken by the Kurmis) is found in the demonstrative system. Most Indo-Aryan languages, including the other Sadani languages, have a two-way distinction between ‘proximal’ and ‘distal’ in demonstrative forms. However, Kurmali makes a three-way distinction between ‘proximal’, ‘distal’ and ‘remote’, *i* ‘DEM.PROX’, *u* ‘DEM.DIST’, and *hau* ‘DEM.REMOTE’. A form similar to *hau* is not found in the neighboring Indo-Aryan languages. It refers to objects which are farther away from the speaker than *u* but which are still visible to the speaker. Consider (28).

- (28) *i=ɬa* *mɔr*, *u=ɬa* *ritu=k*
 DEM.PROX=CLF 1SG.GEN DEM.DIST=CLF PN=GEN
ar *hau=ɬa* *tɔr* *dhiɾha* *lag-ek*.
 and DEM=CLF 2SG.GEN house COP-3SG
 ‘This is my house, that’s Ritu’s and the one even further is yours.’

As in other Sadani languages (except Sadri/Nagpuri and some dialects of Khortha), Kurmali also shows limited morphological ergativity, although the restrictions on its use in Kurmali are somewhat different than in Khortha. In Kurmali, morphological ergativity is limited to transitive subjects expressed by nouns and is found both in present and past tenses, as illustrated in (29) and (30). As our examples suggest, the ergative marker =*ē* is found in stems ending in a consonant or -*a*, while =*ĩ* is found elsewhere.

- (29) *ritu=ĩ* *bhat* *nuɾ-ei(ɔi)*.
 PN=ERG rice eat-PRS.3SG
 ‘Ritu eats rice.’
- (30) *lɔchuwa=ē* *gach=lek* *phɔr* *tɔɾ-l*.
 PN=ERG tree=EXT fruit pluck-PST
 ‘Lachuwa plucked a fruit from the tree.’

As in Khortha, Kurmali also has polypersonal agreement, but unlike in Khortha only subjects and definite objects, both animate and inanimate, are marked on the verb, but not indefinite objects. This is shown in (31)–(33).

- (31) *pasu=ĩ piṭha nuṛ-l-ak.*
 PN=ERG rice.bread eat-PST-3SG
 ‘Pasu ate rice bread.’
- (32) *pasu=ĩ pakhɔn=ʈi=ke phuṭau-l-ei-k.*
 PN=ERG stone=CLF=OBL break-PST-3SG-3SG
 ‘Pasu broke the stone.’
- (33) *ritu=ĩ chɔwa=ʈi=ke thanau-l-ei-k.*
 PN=ERG child=CLF.F=OBL see-PST-3SG-3SG
 ‘Ritu saw (his) daughter.’

With plural objects, the verb only agrees with the number of the object but does not show person agreement. However, it agrees with both person and number of the subject. Consider (34)–(35).

- (34) *mɔĩ tɔhɔra=ke ɖak-l-ɔ(i)-hɔn.*
 1SG 2PL=OBL call-PST-1SG-PL
 ‘I called you (pl.).’
- (35) *mɔĩ ɔkhɔra=ke ʈan-l-ɔ(i)-hɔn.*
 1SG 3PL=OBL pull-PST-1SG-PL
 ‘I pulled them.’

In summary, while Kurmali is clearly a Sadani language, there are also signs that the ancestors of the present-day Kurmis once spoke a different language which was not Indo-Aryan, Dravidian or Munda. Otherwise, it closely resembles Khortha in having ergativity, although with a somewhat different distribution than in Khortha, and a few loanwords from Santali, although fewer than Khortha. As such, with the exception of the unknown “Kurmi language” which has left such a clear mark on the lexicon of modern Kurmali, the same general comments hold here as for Khortha, namely clear signs of contact-induced changes. In Kurmali, this is with respect to the lexicon and possibly also the demonstrative system.

3.4 Panchparganiya

Panchparganiya (also Panchpargania) is referred to as such as it is traditionally spoken in five subdivisions (i.e., *panc pargana*) of the Ranchi district of Jharkhand, namely Bundu, Barendra, the earlier subdivision Sonahatu (now split into the subdivisions Sonahatu and Rahe), Silli, and Tamaṛ (referred to as *Sibubarat* locally). Besides these areas, substantial numbers of speakers are also found in the nearby

pargana Angaṛa. Sonahatu and Rahe now form the core areas of Panchparganiya speakers. In comparison with the other Sadani languages, the number of speakers of this language is relatively small; on the basis of the latest census report (GOI 2011), there are 257,000 Panchparganiya speakers. Alternative names for Panchparganiya include Tamaṛia, Diku Kaji, Gāwari, and Kherwari.

As this language is quite small compared to the other Sadani languages, it is not widely used as a *lingua franca*. However, some of the tribes living in the above-mentioned areas now speak Panchparganiya as their first language. For example, among our Panchparganiya consultants two of the regular informants were from two different ethnic groups, Mundari and Oraon, although Panchparganiya is their native language. Other tribes who speak Panchparganiya as their L1 include the Birhor, Bediya, Patar, Mahali, Lohara, and Chik Baṛaik.

Despite its otherwise strong similarities to Sadri/Nagpuri, there are a number of aspects in which Panchparganiya is closer to Khortha and Kurmali than to Sadri/Nagpuri. For example, the locative marker =*mahan* as in Kurmali (alternatively pronounced as =*mehane*) is also attested in Panchparganiya. Another case in point is plural marking: plurality is marked in Panchparganiya by clitics, such as =*gila* with nouns, and =*man* with third person pronouns and the second person honorific pronoun *raure*. Thus the system of plural marking is somewhat more complex in Panchparganiya than in Sadri/Nagpuri, where =*man* is found in all environments. The classifier =*gāṛa*, borrowed from Munda, is also attested in Panchparganiya, as in Khortha and Kurmali.

There is morphological ergativity in Panchparganiya, which appears to be restricted to focused agents. The ergative suffix =*ē* is obligatory when the object is definite and human, and optional when the object is definite but non-human; it does not occur when the object is indefinite and non-human. It is restricted to nouns, including proper nouns, and is only found in the past tense. (36)–(38) illustrate its use.

- (36) *pandu=ē bhai=ṭa=ke de-l-k.*
 PN=ERG brother=CLF=OBL give-PST-3SG
 ‘Pandu gave (it) to his brother.’

- (37) *andu=ē gach=ṭa=ke kaṭ-l-k.*
 PN=ERG tree=CLF=OBL cut-PST-3SG
 ‘Pandu cut the tree.’ (definite object)

- (38) *pandu pakhan ṭuṭa-l-k.*
 PN stone break-PST-3SG
 ‘Pandu broke stones.’ (indefinite object)

Like Sadri/Nagpuri, there is no polypersonal verb marking in Panchparganiya. This is perhaps due to the fact that it is barely used as a *lingua franca*, hence it would not be expected to incorporate many grammatical features from Munda languages. This also fits in well with the fact that Panchparganiya does not have especially many loanwords from Munda. Taken together, this suggests that Panchparganiya, through its relative isolation (at least in comparison with the other three Sadani languages) is in many ways more conservative than these, both with respect to its lexicon as well as its morphosyntax.

4 Summary and conclusions

The preceding sections have shown that Sadri/Nagpuri, Khortha, Kurmali and Panchparganiya form a clear, compact genealogical Sadani group within the Magadhan sub-group of Indo-Aryan, and all four of the quantitative analyses undertaken, i.e., UPGMA and Neighbor-Joining for both lexical and phonetic similarity, show that all four Sadani languages are more closely related to one another than to any other language in the family, including the other Magadhan languages. This supports the locally held view that these four languages all descend from a single “mother” language which no longer exists in that form, i.e., Sadani, which is why these four languages are referred to collectively as such in the region. This contradicts the classification by most researchers from outside the region, who have tended to view Sadri/Nagpuri as a dialect of Bhojpuri and the other Sadani languages as dialects of Magahi or Maithili, a view which we believe is primarily due to various morphosyntactic features, such as the presence or absence of polypersonal verb marking, etc.

We argue in the preceding sections that it was primarily the respective contact situation of each of these four languages which led to many of the differences between them which we find today. These are summarized in Table 2.

Under the assumption that the earliest Sadani language had morphological ergativity when it was brought to this region, since ergativity is found in three out of four modern Sadani languages, we can say that Sadri/Nagpuri has likely lost ergativity due to its widespread use as a *lingua franca*, which is conducive to “simplification,” especially when none of the contact languages have ergativity. Also, the general restriction to agents expressed by lexical nouns and the past tense in the other three languages – with different extensions and restrictions in each language – suggests that ergative marking was restricted to agentive nouns in the past tense in an earlier stage of Sadani.

However, Sadri/Nagpuri has not only lost complexity, it has also gained complexity in that the distinction between alienable and inalienable attributive

Table 2: Structural similarities and differences between the four Sadani languages.

Feature	Sadri/ Nagpuri	Khortha	Kurmali	Panchparganiya
Large number of loanwords from a small number of donor languages	–	from Santali	from an unknown source likely unrelated to Indo-Aryan, Dravidian or Munda; somewhat less also from Santali	–
Ergativity	–	agent is a noun or 1st person, singular; only in the perfective	agent is a noun; spread from use only with the past to the present tense as well	only found with focused agents, only when the agent is a noun and only in the past tense
Polypersonal marking	–	only 1→3, extends to addressee marking	only with definite objects	–
Alienable/inalienable distinction in attributive possession	only with 3rd person possessors	–	–	–
Narrative category	only in the present tense	–	–	–

possession, found in all Munda languages, both north and south, is now found in this language, albeit only with a third person possessor. Also, through the loss of the vowel /a/ under certain conditions, a new verbal category has been created in the present tense in Sadri, namely the narrative, which closely mirrors the distribution of the homophonous finite marker in North Munda. Finally, Sadri shows no high number of loanwords from a single language, as is to be expected in such a multilingual environment. Instead it itself serves as the donor language for countless loans in its contact languages.

In contrast, Khortha has quite a large number of borrowings from Santali, its primary contact language, while Kurmali has several prominent loans from an unknown language which no longer appears to be spoken, as well as a smaller number of Santali loanwords, its primary contact language at present. Otherwise, Khortha and Kurmali are quite conservative, as both retain morphological ergativity

but have not developed an alienable/inalienable distinction in attributive possession nor a narrative category.

Panchparganiya, the Sadani language with the least amount of contact of all, is unsurprisingly also the most conservative: It does not have a noticeably large number of loans from a single donor language, it has retained morphological ergativity, and it has not developed an alienable/inalienable distinction in attributive possession, nor a category similar to the narrative in Sadri/Nagpuri.

With respect to polypersonal marking on the verb, it is not entirely clear whether this type of marking was already present when Sadani was first brought to Jharkhand, so that we cannot say with certainty whether it represents an innovation or a retention from older stages of Sadani. Within Indo-Aryan it is largely, if not entirely, restricted to Magadhan, but it is not found universally within this group, even outside of Sadani. For example, while it is found in several non-Sadani Magadhan languages, such as Maithili, Magahi, etc., as well as Khortha and Kurmali, it is not found e.g. in Bhojpuri, nor in Sadri/Nagpuri or Panchparganiya. However, the distribution of this type of marking within Sadani may give us some clues as to its origin.

The lack of polypersonal marking in a *lingua franca* such as Sadri/Nagpuri is hardly surprising, as it is used by speakers of several different L1's, many of which do have polypersonal marking, but some of which do not. However, the lack of this type of marking in Panchparganiya as well, which is otherwise quite conservative with respect to morphosyntax, suggests that polypersonal marking is an innovation in Khortha and Kurmali.

In this respect it is important to note that Khortha and Kurmali are the two languages with contact largely restricted to one major North Munda language, Santali, where polypersonal marking is very productive, and that to the north of these, directly bordering the Khortha-speaking region, Magahi is spoken, where polypersonal marking is also found. It seems likely that it is precisely this feature, which Khortha (and Kurmali) shares with Magahi, which has led many researchers to consider Khortha to be a dialect of Magahi instead of belonging to an independent Sadani group.

Thus, polypersonal marking in Magadhan languages appears to be an areal feature, primarily found in the eastern Magadhan languages, including Khortha and Kurmali, and seems to be a direct result of intense contact between the Magadhan languages and North Munda, from where it has spread to other languages.¹³ Further work will hopefully shed more light on this issue.

¹³ This is not to say that polypersonal marking is entirely restricted to Magadhan languages within Indo-Aryan, only that it seems to be restricted to these languages within our region of study, i.e., the eastern Indo-Gangetic Plain.

In summary, a comparison of all five features investigated here for all four Sadani languages reveals much about earlier stages of Sadani, namely that:

- It possessed ergativity, which was likely restricted to agents expressed by nouns and found only in the past tense. From this origin it spread in different languages to include e.g. the first person singular in Khortha, but only in the perfective, and also to the present tense in Kurmali, but still with nouns only. In Panchparganiya it is still restricted to nouns and the past tense, but also only with focused agents;
- polypersonal marking most likely represents an innovation in Khortha and Kurmali, i.e., those languages with one primary contact language, namely a North Munda language – a tendency which is reinforced by areal pressure;
- other innovations, such as a distinction between alienable and inalienable attributive possession and a productive narrative category, are only found in Sadri, due to its extensive use as a *lingua franca*;
- Panchparganiya on the other hand, historically somewhat of a hermit with respect to contact, is also the most conservative with respect to its morpho-syntax and lexicon.

A number of questions still remain. For example, how did a typical North Munda feature such as the distribution of /a/ in Sadri/Nagpuri not also arise in Khortha and Kurmali, the two Sadani languages most closely in contact with primarily a single North Munda language – whereas it did in Sadri/Nagpuri, which also has close contact with Kharia (South Munda) and Dravidian languages, none of which has this trait? Instead, Sadri/Nagpuri alone has exploited the presence vs. absence of the final /a/ in the present tense to express a modal/pragmatic category which neither of the other two languages did. A possible answer lies in the fact that the present tense currently found in these two languages evolved from an entirely different, periphrastic construction, likely an older progressive category, so that there likely never was an/a/ or /la/ in the present tense in Khortha and Kurmali, i.e., a distinction of this type could not have become productive there. However, further work is required to confirm this.

In fact, much more work is necessary on all four Sadani languages and on the Munda and Dravidian languages of this region as well, as we still know virtually nothing of the differences in language use among the different castes and ethnic groups in each language, nor about subtle, regionally based language differences. To give just one example, we do not know if certain ethnic groups whose traditional L1 lacks ergativity use ergative marking differently than most other speakers of Khortha, Kurmali or Panchparganiya, etc. It is hoped that further research will provide further answers to these and other questions in the near future, although it is also likely to raise new, interesting questions as well.

Abbreviations

1,2,3	Person
ABL	ablative
ACT	active
ADDR	addressee
ALL	allative
AMB	ambulative
ANTIC	anticipatory
APPL	applicative
AUX	auxiliary
CLF	classifier
CONT	continuative
COP	copula
CVB	converb
DEM	demonstrative
DIST	distal
DU	dual
ERG	ergative
EXT	extensional
F	feminine
FIN	finite marker
FOC	focus
FUT	future
GEN	genitive
IMP	imperative
IND	indicative
INF	infinitive
INS	instrumental
LNK	linker
LOC	locative
NAR	narrative marker
NP	nominal phrase
OBJ	object
OBL	oblique
PRF	perfect
PL	plural
PN	personal name
PROX	proximate
PRS	present
POSS	possessive
PST	past
SEQ	sequential converb
SG	singular

SUBJ	subjunctive
TOP	topic
TR	transitive
V2	'vector verb'

Acknowledgment: Research for this article was funded by the German Research Council (DFG) project “Towards a linguistic prehistory of eastern central South Asia (and beyond)”, Project Grant 326697274, which we gratefully acknowledge here.

Research funding: This research was funded by a generous grant by the German Research Council (Deutsche Forschungsgemeinschaft, DFG, grant no. 326697274).

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.
- Anderson, Gregory D. S. 2007. *The Munda verb. Typological perspectives*. Berlin & New York: Mouton de Gruyter.
- Bloch, Jules. 1919. *La formation de la langue marathe*. Paris: Champion.
- Campbell, Andrew. 1899. *A Santali–English dictionary*. Pokhuria Manbhum: Santal Mission Press.
- Chatterji, Suniti Kumar. 2002 [1926]. *The origin and development of the Bengali language*. Delhi: Rupa.
- COG. 2019. Cog. Version (1.3.4.10016). SIL International. <http://software.sil.org/cog/> (accessed 9 July 2019).
- Eberhard, David M., Gary F. Simons & Charles D. Fennig (eds.). 2019. *Ethnologue: Languages of the world*. Dallas, Texas: SIL International.
- GOI [Government of India]. 2011. *Census of India 2011*. New Delhi: Office of the Registrar General and Census Commissioner, India, Ministry of Home Affairs.
- 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, Government Printing.
- Jordan-Horstmann, Monika. 1969. *Sadani. A Bhojpuri dialect spoken in Chotanagpur*. Wiesbaden: Otto Harrassowitz.
- JTWRI [Jharkhand Tribal Welfare Research Institute]. 2013. *Language diversity in Jharkhand. A study on socio-linguistic pattern and its impact on children's learning in Jharkhand*. Morabadi, Ranchi: JTWRI.
- Kobayashi, Masato. 2012. *Texts and grammar of Malto*. Vizianagaram: Kotoba Systems.
- Kobayashi, Masato & Tirkey Bablu. 2017. *The Kurux language. Grammar, texts and lexicon*. Leiden & Boston: Brill.
- McGregor, Ronald S. (ed.). 1997. *The Oxford Hindi–English dictionary*. Oxford & Delhi: Oxford University Press.
- Neukom, Lukas. 2001. *Santali*. Munich: Lincom Europa.
- Nichols, Johanna & Warnow Tandy. 2008. Tutorial on computational linguistic phylogeny. *Language and Linguistics Compass* 2(5). 760–820.

- Nowrangi, Peter Shanti. 1956. *A simple Sadānī grammar*. Ranchi: D.S.S. Book Depot.
- Osada, Toshiki. 2008. Mundari. In Gregory D. S. Anderson (ed.), *The Munda languages*, 99–164. London & New York: Routledge.
- Paudyal, Netra P. In progress. *A grammar of Khortha*.
- Peterson, John. 2015. From “finite” to “narrative” – the enclitic marker =a in Kherwarian (North Munda) and Sadri (Indo-Aryan). *Journal of South Asian Languages and Linguistics* 2(2). 185–214.
- Peterson, John. 2017. Fitting the pieces together – towards a linguistic prehistory of eastern-central South Asia (and beyond). *Journal of South Asian Languages and Linguistics* 4(2). 211–257.
- Peterson, John & Sunil Chik Baraik. In press. *A grammar of Sadri/Nagpuri – An Indo-Aryan lingua franca of eastern central India*. Mysore: Central Institute of Indian Languages (CIIL).
- Tiwari, Udai Narain. 1960. *The origin and development of Bhojpuri*. Calcutta: The Asiatic Society.
- Yadav, Dev Narayan. 2012. *Sadri grammar: A sketch*. Lalitpur: Sajha Prakashan.