John Peterson*

Introduction – advances in the study of Munda languages

DOI 10.1515/jsall-2015-0008

South Asia is home to around 500 languages belonging to at least six different genetic stocks: Indo-Aryan, a branch of the Indo-European phylum, Dravidian, Andamanese,¹ Tibeto-Burman, Tai, and Austro-Asiatic, as well as a number of language isolates, including Burushaski, Kusunda, and most likely also Nihali, although there is some debate on the classification of this language.²

The Munda languages, which are the main topic of the present issue, make up the western branch of the Austro-Asiatic phylum and stretch from the central Indian states of Maharashtra and Madhya Pradesh in the west through to eastern Nepal and western Bangladesh in the east, being concentrated primarily in the states of Jharkhand, Odisha (formerly known as “Orissa”), and neighboring regions in the states of Chhattisgarh, Andhra Pradesh and West Bengal. Figure 1, from Anderson (2007: 7), presents a general overview of the regions where the various Munda languages are spoken.

The eastern branch of Austro-Asiatic, stretching roughly from the northeastern Indian state of Meghalaya in the northwest and the Nicobar Islands in the southwest through to Vietnam in the east, is generally referred to as Mon-Khmer. According to Lewis et al. (2015), 169 languages in total belong to the Austro-Asiatic phylum, of which 146 belong to the Mon-Khmer branch and 23 to the Munda branch.

There is general consensus that the Munda group can easily be divided into two clearly distinguishable branches, namely North Munda and South Munda. Beyond this very basic division, however, there continues to be some debate over further details, especially with respect to the internal classification of South Munda. Although we cannot go into detail here on these various classifications and the logic behind them, in the following we present a number of genetic

1 Abbi (2009) has recently argued that what is commonly viewed as the “Andamanese” family in fact consists of two genetically unrelated groups, whose protolanguages she refers to as “Proto Ang” and “Proto Great Andamanese”, so that the exact number of language families in South Asia is still an open issue.

2 Cf. Zide (2008b) for a detailed discussion on the status of Nihali.

*Corresponding author: John Peterson, Department of General Linguistics, University of Kiel, Germany. E-mail: jpeterson@isfas.uni-kiel.de
classifications which have been proposed for the Munda languages and briefly
discuss how these models differ from one another.

A (greatly simplified) version of the classification of Austro-Asiatic based on
the discussion in Pinnow's (1959: 1–3) landmark study of the historical phonol-
ogy of the South Munda language Kharia – in actual fact a historical study of
Munda (and beyond) in general – is presented in Figure 2. For ease of presenta-
tion, the members of the “Eastern Group” (Ost-Obergruppe), which we refer to
here collectively as “Mon-Khmer”, are not presented here.

Ten years later, Zide (1969: 412) presented a somewhat different classifica-
tion of this group. The first major difference to Pinnow's (1959) classification is
that the latter’s “Eastern Subgroup” and “Western Subgroup” of Munda have
been combined under the term “North Munda” by Zide, with Korku now forming
a separate branch of “North Munda”. Furthermore, the languages of Pinnow’s
“Eastern Subgroup”, which Zide refers to as “Kherwarian”, now show a further

Figure 1: Map of the Munda languages (from Anderson 2007: 7, reprinted with kind permission by Mouton de Gruyter).
internal classification in Zide’s scheme, instead of being on a more or less equal footing with no further internal classification in Pinnow’s classification: In Zide’s scheme, Santali is now in a separate group from the remaining Kherwarian languages ("Mundari-Ho" in Zide’s terminology). Furthermore, Pinnow’s “Central Subgroup” and the “Southern Subgroup” now together form the “South Munda” branch in Zide’s (1969) scheme, i.e., Kharia and Juang here form a branch of South Munda instead of branching off directly from the Munda group, as in Pinnow (1959).\(^3\) Pinnow’s “Southern Subgroup” now forms the “Koraput Munda” branch in Zide’s (1969) scheme, which then divides into “Sora-Gorum” and “Gutob-Remo-Gtaʔ”. This is presented in Figure 3.

Zide’s (1969) classification has become the most widely accepted and is now generally viewed as the “traditional” classification of the Munda languages (cf., e.g., Anderson 2008a: 2), although it, too, has recently been the subject of further discussion. One important revision to this classificatory scheme is Anderson (1999), in which South Munda is directly divided into three daughter groups, namely Sora-Gorum, Kharia-Juang and Gutob-Remo-Gtaʔ, i.e., without

---

\(^3\) Cf. Zide and Stampe (1968) on the place of Kharia and Juang in Munda.
the “Koraput Munda” group assumed in Zide (1969). This is illustrated in Figure 4.

One last classification should be mentioned here, namely that of Anderson (2001) for South Munda, which differs in a number of important ways from the model proposed in Anderson (1999). In this model, given below in Figure 5, Proto-Sora-Gorum separated at a rather early stage from the rest of South Munda, leaving behind a continuum of groups, consisting of Proto-Juang, Proto-Kharia, Proto-Gutob-Remo and Proto-Gtaʔ.

The eastern branch of Austro-Asiatic, i.e., the Mon-Khmer group, is not only far larger than the Munda group in terms of the number of languages; it also
shows considerably more internal complexity than the latter, western group. As the present issue is primarily dedicated to the Munda languages, we will restrict ourselves here to a simplified version of the internal classification of Mon-Khmer given in Lewis et al. (2015), presented here in Figures 6(a)–6(f), as several of these subgroups will be mentioned in Anderson (this issue, see below). Note that the internal structure is only given here for those groups which are mentioned in Anderson’s study.

Serious study of the Munda languages and their speakers began already in the nineteenth century, and many of these early studies are still considered standard reference works. As Anderson (2007: 8–10) presents a general overview of the history of the study of the Munda languages from the nineteenth century onwards, the present discussion is very brief and concentrates on more recent developments in this field and only touches upon the larger and more influential studies before the twenty-first century, making no claim to exhaustiveness.4

First and foremost among these early works are those by Hoffmann and his associates on the North Munda language Mundari, such as the 16-volume Encyclopaedia Mundarica (Hoffmann and van Emelen 1928–1978 [1990]), so that Mundari is arguably the best described Munda language to date with respect to its lexicon. This language is also one of the best described in terms of its morpho-syntax and phonology, thanks again in no small measure to Hoffmann’s grammatical studies of this language (1903, 1905/1909 [2001]). Further, more recent influential studies on this language include Munda (1988), and Osada (1992), among others.

The North Munda language Santali, the largest of all Munda languages in terms of speakers, was also the object of serious documentation at an early date, cf., e.g., Boding’s many works on this language, including his Materials for a Santali Grammar (1922, 1929a), his beginners’ grammar (1929b), but especially

---

4 A detailed bibliography of the individual Munda languages can be found under http://www.isfas.uni-kiel.de/de/linguistik/forschung/southasiabibliography
his five-volume Santali-English dictionary (1929–1936) as well as his invaluable collection of folktales (1925–1929 [1997]), which even today serves as a major source of data for grammatical descriptions of this language, such as Neukom (2001). Other influential works on this language include Campbell and Macphail (1933), Macphail (1964), and Ghosh (1994). The third largest North Munda language, Ho, has also been well documented, most notably by Deeney (1975, 1978) and, more recently, Pucilowski (2013).

In terms of their documentation, the South Munda languages did not fare as well as the North Munda languages before the 1960’s, although there are a number of notable exceptions, such as Bannerjee’s (1894) introduction to Kharia or Archer’s (1942) collection of Kharia songs, although these cannot

---

**Figure 6:** (a) Mon-Khmer upper-level subgroupings. (b) Aslian subgroupings. (c) Eastern Mon-Khmer subgroupings. (d) Northern Mon-Khmer subgroupings. (e) Palyu languages. (f) Viet-Muong subgroupings (based on Lewis et al. 2015).
compare with the studies on North Munda described above.\(^5\) Another South Munda language which received considerable study before the 1960’s was Sora, e.g., studies such as Ramamurti (1931), a grammar of the Sora language, and Ramamurti (1938 [1986]), a Sora-English dictionary.

The 1960s were a major turning point in Munda linguistics, with a tremendous upsurge in the number of studies on these languages. The first major study to be mentioned here is Pinnnow (1959), the author’s “Habilitationsschrift” or “professorial dissertation”, dedicated primarily to the historical phonology of Kharia, although in fact this study contains a wealth of data demonstrating the historical relations between the Munda languages in general and also the relations of this group to its distantly related Mon-Khmer cousins. Although now clearly dated, this study is still mandatory reading for all interested in Munda (and Austro-Asiatic) historical linguistics. Other influential works by this same author include Pinnnow (1965a), a Kharia text collection from spoken narratives, written texts, etc., as well as a further, smaller study which appeared in the same year (1965b). Pinnnow also composed numerous other studies which are still standard reading for all interested in the field of Munda languages, most importantly Pinnnow (1966), a comparative study of the Munda verb. Other notable studies on Kharia from this time include Biligiri’s (1965) grammar and lexicon and, somewhat later, Malhotra’s (1982) grammar of the same language, both of which were the respective author’s PhD dissertation.

Another hub of research on these languages outside of India during this period was in the United States, especially the project under the auspices of Norman Zide at the University of Chicago to document these languages. Numerous studies – far too many to mention here – appeared during this project which continue to have an enormous impact on the field, two of which were mentioned above, Zide (1969) and Zide and Stampe (1968). Important work on the Munda languages was also taking place elsewhere in the United States in conjunction with this project, e.g., Fernandez’ (1967) grammar of Remo, and Matson’s (1964) grammatical sketch of Juang, both of which were the respective author’s PhD dissertation.

Much work on the Munda languages has also traditionally been conducted in India itself, a number of these studies having already been mentioned above, such as Banerjee (1894), Biligiri (1965), Ghosh (1994), Malhotra (1982), and Munda (1988), but also Bhattacharya (1968), a dictionary of Bonda (Remo), and many others, and an indigenous scholarly tradition has now developed at

\(^5\) Non-linguistic studies, such as Roy and Roy’s (1937) monumental ethnological study of the Kharia, will not be dealt with here, despite their importance for the field of Munda studies, as they are not primarily concerned with the languages of these ethnic groups.
least for the major Munda languages, e.g. for Kharia (J. Đunque 1999, M. Đunque 1986; Kullü 1981, 2000, etc.) and also for Santali, Mundari and Ho.

With the beginning of the twenty-first century a further burst of research activity on the Munda languages began which is still in full swing. The most important of these recent studies is without a doubt Anderson (ed.) (2008) with its sketch grammars of the various Munda languages – Anderson (2008b) on Gtaʔ, Anderson and Harrison (2008a) on Sora, Anderson and Harrison (2008b) on Remo/Bonda, Anderson and Rau (2008) on Gorum, Anderson et al. (2008) on Ho and other Kherwarian languages, Ghosh (2008) on Santali, Griffths (2008) on Gutob, Kobayashi and Murmu (2008) on Keraʔ Mundari, Osada (2008) on Mundari, Patnaik (2008) on Juang, Peterson (2008) on Kharia and Zide (2008a) on Korku. As many of these languages are endangered and since many had only received rudimentary documentation prior to the appearance of this volume, the value of this volume for the field of Munda linguistics cannot be overstated. Further recent descriptive studies include Minegishi and Murmu (2001), an annotated Santali lexicon, Rau (2011) on grammatical voice in Gorum, Peterson (2006), a ca. 900-page work on Kharia, including a full grammatical description of the language (since published as Peterson 2011a), a collection of segmented, glossed, annotated and translated texts (since published as Peterson 2011b) as well as a Kharia-English lexicon (since published as Peterson 2009), and Peterson (2011c) on the status of phonological, morphosyntactic and orthographical “words” in Kharia.

The Munda languages are also playing an increasingly important role in typological and theoretical linguistics. For example, regardless of the position one takes on the universality of parts of speech such as “noun”, “verb” and “adjective” or how these are to be defined, the Munda languages now feature prominently in any typological discussion of parts of speech. Cf. e.g. Evans and Osada (2005a) and the invited responses to this article (Peterson 2005; Hengeveld and Rijkhoff 2005; Croft 2005) as well as the authors’ reply to these (Evans et al. 2005b). Further contributions in this area include Peterson (2007) and the articles dealing with Munda languages in Rijkhoff and van Lier’s (2013) multi-authored volume on languages with flexible word classes (Don and van Lier 2013; Peterson 2013; Rau 2013). Other recent typological studies on Munda include Anderson’s (2007) detailed, book-length study of the verb in Munda languages, Donegan and Stampe (2004), which argues that many common features of Munda and Indo-Aryan are prosodically motivated, Leufkens (2011) on transparency in Kharia, Peterson (2003) on finiteness in Kharia, Peterson (2014), dealing with “number” in Kharia from a typological and sociolinguistic perspective (making use of Silverstein’s (2003) concept of “indexicality”), works on areal linguistics involving Munda (Peterson 2010; in press), Peterson (2011d)
on Kharia within the framework of Role and Reference Grammar (RRG), and Peterson and Maas (2009) on reduplication in Kharia.

The present issue adds three new, original contributions to this growing list. In the first of these, Anderson (this issue), the author focuses on the phonological and prosodic characteristics of Munda words and phrases, and puts this data into the larger, Austro-Asiatic perspective. Among other things, Anderson calls attention to parallels between Mon-Khmer languages with their sesquisyllabic or “weak-strong” prosodic word structure and the typical iambic structure of phonological words in Munda languages. He also notes that while the free-standing form of nouns is generally not reconstructible for Munda, the nominal root typically is, suggesting that once productive derivational strategies involving prefixation, infixation, suffixation or reduplication — or in some cases compounding — can be assumed from a historical perspective for both Munda and Mon-Khmer in order to fulfill the bimoraic constraint on free-standing forms which is so typical for Austro-Asiatic languages in general. This results in the present-day situation of clearly related forms which cannot be reconstructed to a single, common word form at an earlier stage, while at the same time providing yet another convincing demonstration of the unity of the Austro-Asiatic phylum as a whole.

The second contribution, Peterson (this issue), is a continuation of earlier studies by this author into Munda / Indo-Aryan language contact in eastern-central South Asia. Here Peterson takes a closer look at the marker of the indicative present tense in Sadri, the Indo-Aryan lingua franca in much of central and western Jharkhand state for speakers of North Munda, Kharia (South Munda) and several smaller Indo-Aryan languages. This marker derives historically from the category which is now the subjunctive but which was previously the indicative present tense plus * = lā, which was once presumably an enclitic focus marker, a development found in many Indo-Aryan languages. In Sadri however, unlike in most other Indo-Aryan languages where this development took place, this marker has been re-analyzed as -l, which now forms part of the suffix marking the indicative present, and the “optional” enclitic marker = a. After analyzing the distribution of = a and its apparent functions in Sadri narratives collected by the author, Peterson argues that = a has come to function similarly to the homophonous “finite marker” = a in Kherwarian languages, presumably through contact with these speakers using Sadri as a lingua franca. The article then shows how this marker serves to further define an area of linguistic convergence between Indo-Aryan and Munda.

In the third and final contribution, Voß (this issue), Judith Voß takes a closer look at “person marking” in its broadest sense in the South Munda language Gutob, based on data she collected during fieldwork in the village of
Jalarhanjar in the Koraput District of the Indian state of Odisha (Orissa). As Voß notes, there are three positions for person markers in Gutob: sentence-initial position, postverbal position, or enclitic to a preverbal element, generally – although not always – the element which directly precedes the clause-final verb. In her study, Voß analyzes these markers in all three positions with respect to their status as independent pronouns and as agreement markers. In accordance with earlier research, she comes to the conclusion that sentence-initial forms are best viewed as independent pronouns, while postverbal forms can be considered agreement markers, although she argues that the preverbal enclitic markers do not neatly fit into either category, as they share a number of properties with both of these categories.

These three contributions thus represent important advancements in three different fields of Munda linguistics, namely historical, areal and descriptive linguistics, and provide us with new insights into the individual languages, the influence that the Munda languages exert upon their non-Munda linguistic neighbors, and the ways in which they have diverged from the other members of their own family, while at the same time also retaining many interesting Austro-Asiatic family traits.

References


Peterson, John. this volume. From “finite” to “narrative”: The enclitic marker =a in Kherwarian (North Munda) and Sadri (Indo-Aryan).


Voß, Judith. this volume. Person markers in Gutob.


