Re-evaluating the linguistic prehistory of South Asia

Originally presented at the workshop: Landscape, demography and subsistence in prehistoric India: exploratory workshop on the middle Ganges and the Vindhyas

Leverhulme Centre for Human Evolutionary Studies
University of Cambridge, 2-3 June, 2007

Published in; Toshiki OSADA and Akinori UESUGI eds. 2008. Occasional Paper 3: Linguistics, Archaeology and the Human Past. pp. 159-178. Kyoto: Indus Project, Research Institute for Humanity and Nature.

Roger Blench
Kay Williamson Educational Foundation
8, Guest Road
Cambridge CB1 2AL
United Kingdom
Voice/ Fax. 0044-(0)1223-560687
Mobile worldwide (00-44)-(0)7967-696804
E-mail R.Blench@odi.org.uk
http://www.rogerblench.info/RBOP.htm

Re-evaluating the linguistic prehistory of South Asia

Roger Blench
Mallam Dendo Ltd
E-mail R.Blench@odi.org.uk
http://www.rogerblench.info/RBOP.htm

ABSTRACT

South Asia represents a major region of linguistic complexity, encompassing at least five phyla that have been interacting over millennia. Although the larger languages are well-documented, many others are little-known. A significant issue in the analysis of the linguistic history of the region is the extent to which agriculture is relevant to the expansion of the individual phyla. The paper reviews recent evidence for correlations between the major language phyla and archaeology. It identifies four language isolates, Burushaski, Kusunda, Nihali and Shom Pen and proposes that these are witnesses from a period when linguistic diversity was significantly greater. Appendices present the agricultural vocabulary from the first three language isolates, to establish its likely origin. The innovative nature of Kusunda lexemes argue that these people were not hunter-gatherers who have turned to agriculture, but rather former cultivators who reverted to foraging. The paper concludes with a call to research agricultural and environmental terminology for a greater range of minority languages.

INTRODUCTION

BACKGROUND

The world's languages can be divided into phyla and language isolates. A language phylum is a genetic grouping of languages not demonstrably related to any other, typically Austronesian or Indo-European. Language isolates are individual languages or dialect clusters that have not been shown to be related to other languages. Most of the world is occupied by populations speaking a fairly restricted number of language phyla, which suggests that these languages have spread (either by actual movements of population or by assimilation of other languages) in fairly recent times. There is a broad relationship between the internal diversity of a language phylum and its age. For example, if indeed the languages of Australia can all be related to one another, the proto-language must go back to the early settlement of the continent, to account for their high degree of lexical diversity. However, through most of the world, the coherence of phyla or their branches is more transparent than in Australia, and as a consequence we can ask what engine drove the dispersal of a particular language grouping and can this be detected through correlations with the methods of other disciplines, notably archaeology and genetics? For a language phylum such as Austronesian, with its dispersal from island to island, and broadly forward movement, this type of approach has been particularly successful. Elsewhere on the linguistic landscape, the results are more controversial, in part because of lacunae in the data but also the nature of research traditions. This paper [1)] looks at the language situation in South Asia and the regional potential for this type of interdisciplinary reconstruction of prehistory.

METHODOLOGICAL ISSUES

Linguists typically use the comparative method to identify language phyla, comparing as many candidate languages as possible, trying to identify common features of phonology, morphology and lexicon and

excluding those which do not meet these criteria (Durie and Ross 1996). Most of the language phyla of the world have no written attestations, and as a consequence hypotheses must draw entirely upon modern (i.e. synchronic) language descriptions. One methodological consequence of this is that all languages are treated as of equal importance; indeed moribund languages or those with small numbers of speakers may well be crucial to historical reconstruction.

However, where early written forms exist, historical linguists can be seduced into forgetting these principles in favour of privileging written forms. Sanskrit, Old Chinese and Old Tamil are typically considered representations of the proto-form of a language family. Hence Turner's (1966) compilation of Comparative Indo-Aryan begins with Sanskrit and seeks modern reflexes of the attested forms, rather than reconstructing proto-forms from modern languages and searching Sanskrit for cognates. Similarly, the Burrow and Emeneau (1984) Dravidian Etymological Dictionary is centred on Tamil. Wholly unwritten phyla such as Niger-Congo and Austronesian proceed in a quite different manner, deducing common roots from comparative wordlists of modern languages and thereby moving to apical reconstructions. Indo-Aryan and Dravidian thus have 'common forms' but not historical reconstructions, because typically, the compilers of etymological dictionaries do not clearly develop criteria for loanwords as opposed to true reflexes. Southworth (2006) also points out that semantic reconstructions tend to focus on meanings in written languages, which may be remote from the actual referent in the proto-language[2)].

A consequence is that for phyla where there are significant early written attestations there is a tendency to divide languages into 'major' and 'minor', and to downplay the importance of field research on 'minor' languages. Despite the very large number of Indo-Aryan languages in South Asia (Table 2), new descriptive studies are very sparse, in part because this is a low priority for researchers; reference works (e.g. Masica 1991) simply assume that Indo-Aryan is a demonstrated genetic grouping.

LANGUAGE SITUATION IN SOUTH ASIA

LANGUAGE PHYLA

South Asia is home to number of distinct language phyla as well as language isolates, i.e. individual languages which have no clear affiliation. Often these are thought to be residual, i.e. to be the remaining traces of language families that once existed. The major language phyla of South Asia are shown in Table1.

Table 1 Language Phyla of South Asia

Phylum	Examples
Indo-European	Sanskrit, Hindi, Bengali, Assamese
Dravidian	Tamil, Telugu, Malayalam
Austroasiatic	Munda, Nicobarese, Khasian
Tibeto-Burman	Naga, Dzongkha, Gongduk
Daic	Aiton, Phake
Andamanese	Onge, Great Andamanese

There are also the following language isolates;

Burushaski	(Pakistan)
Kusunda	(Nepal)
Nihali	(India)
Shom Pen	(India)

The following languages are listed as unclassified (Ethnologue 2005), presumably for lack of information.

Aariya, Andh, Bhatola, Majhwar, Mukha Dora, Pao

The Wanniya-laeto (Vedda) in Sri Lanka evidently had a distinctive speech, but it is gone and the fragmentary evidence suggests no obvious affiliation (see review in Van Driem 2001)[3)].

Apart from these, there are languages which seem

very remote from their putative genetic congeners. The Gongduk language of Bhutan seems to be very distinct from other branches of Tibeto-Burman (Van Driem 2001). This may be because it is 'really' a relic of a former language phylum which has been assimilated or because it is an early branching of Tibeto-Burman. No mention of this language is made in recent reference books such as Matisoff (2003) or Thurgood and LaPolla (2003) presumably due to its inconvenient nature.

Tanle 2 shows the numbers of languages by phylum in South Asia (defined as Pakistan, India, Nepal, Bhutan, Bangla Desh, Sri Lanka and the Maldives).

Table 2 Language numbers in South Asia by phylum

Phylum	No. of languages
Andamanese	13
Austroasiatic	6(N)+3(K)+22(M)
Daic	5
Dravidian	73
Indo-Aryan	253
Tibeto-Burman	230
Unknown	6
Isolates	3

Source: Ethnologue (2005)

The calculations exclude external vehicular languages and creoles.

Table 3 shows the numbers of languages spoken in South Asia by country.

Table 3 Language numbers in South Asia by country

Country	No. of languages
Bangla Desh	39
Bhutan	24
India	415
Maldives	1
Nepal	123
Pakistan	72
Sri Lanka	2

Source: Ethnologue (2005)

DOCUMENTATION AND RESEARCH

South Asian languages are documented in a very patchy way; major literary languages are very well known, with large dictionaries that are increasingly online. 'Minor', unwritten, languages are some of the least-studied in the world, due to research restrictions on 'tribals' in India and regrettably, local publications are of highly uncertain quality in India. Linguistic description is related to ideology and nationalism in a very unhealthy way. Pakistan, Nepal, Bhutan are paradoxically much better covered, due to external research input. Unfortunately, recent civil disorder has made research conditions problematic, but nonetheless, a relatively small country like Nepal is much better known than India. Bangla Desh appears as an almost complete blank on the linguistic map.

DRAVIDIAN

The Dravidian languages, of which Tamil is the most well-known, are spoken principally today in south-central India, although Malto and Kurux are found in northeast India and Nepal and Brahui in Pakistan and Afghanistan. Dravidian languages were first recognized as an independent family in 1816 by Francis Ellis, but the term Dravidian was first used by Robert Caldwell (1856), who adopted the Sanskrit word dravida (which historically meant Tamil). Dravidian languages are often referred to as 'Elamo-Dravidian' in modern reference books, especially those focussing on archaeology. As early as 1856, Caldwell argued for a relation between 'Scythian', a bundle of languages that included the ancient language of Iran, Elamitic, and the modern-day Dravidian languages. This argument was developed by McAlpin (1981) and has gained acceptance rather in excess of its true evidential value.

Another worrying subtheme in Dravidian studies is the putative connection with African languages. Although an early idea clearly based on a crypto-racial hypothesis, it is being newly promulgated in the International Journal of Dravidian Linguistics (cf. for example, Winters 2001). The early presence of African crops in northwest India is being seen as proof of a tortuous model that has Mande speakers leaving Africa to spread civilisation across the world (the

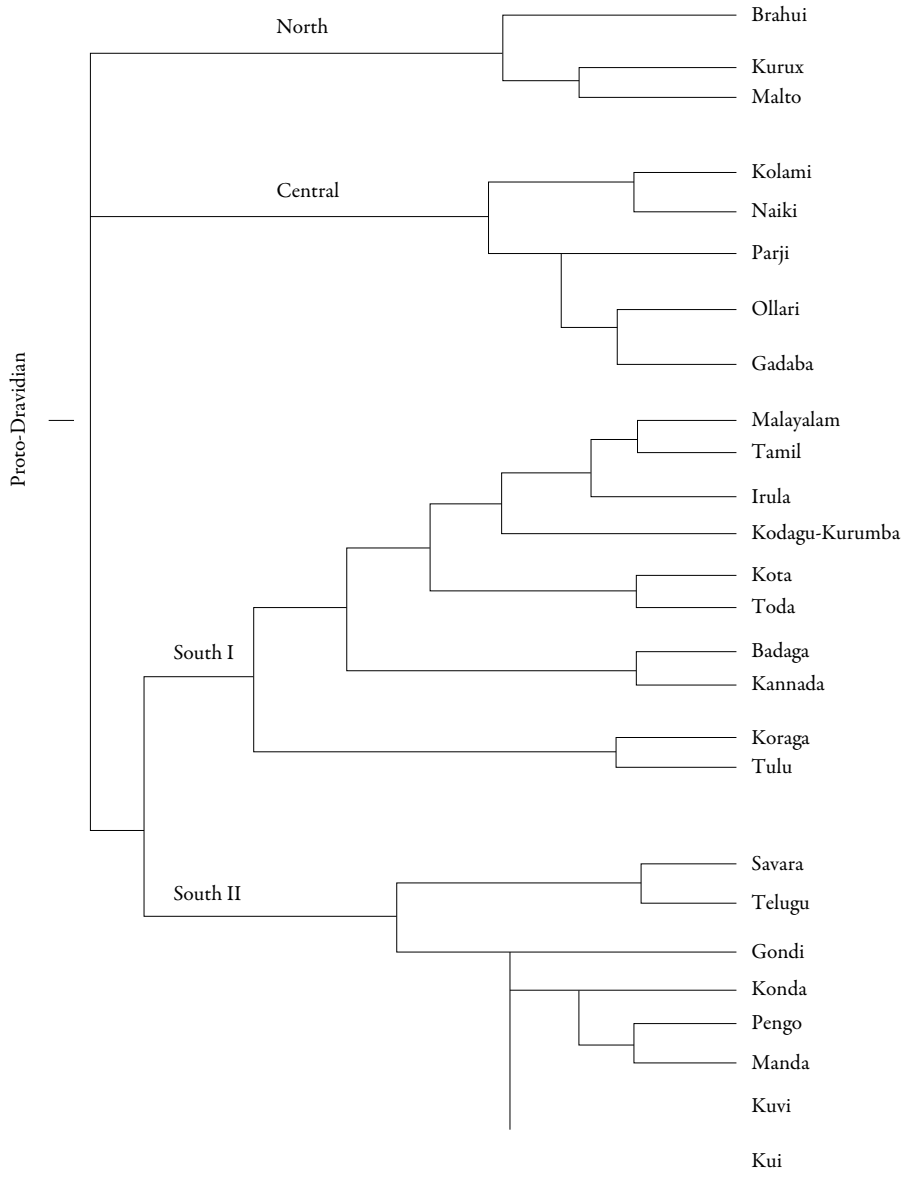

Figure 1 Classification of the Dravidian languages

New World also features in this theory). It should be emphasised that there is no linguistic evidence of any credibility that supports such an unlikely migration.

Surprisingly for a well-known and much-researched group, there are a large number of languages whose Dravidian status is uncertain (see list in Steever 1998: 1) as well as 'dialects' that may well turn out to be distinct languages. Curiously, the standard reference on Dravidian (Krishnamurti 2003: 19) claims that there are only twenty-six Dravidian languages although the Ethnologue (2005) lists 73. Although some of these are dialects of recognised groups, a list of unclassified languages for which almost no published data exists argues that this topic is strewn with uncertainties.

Dravidian divides into either three groups (Zvelebil 1997; Krishnamurti 2003: 21) or four (Steever 1998) since Zvelebil amalgamates Steever's two Southern groups. Figure 1 shows a tree of Dravidian based on these recent classifications.

The presence of North Dravidian languages, particularly Brahui in Pakistan, and the putative link with Elamite has confused much previous thinking about this phylum, with models trying to make proto-Dravidian come from the Near East, and be responsible for the Harappan script, etc. But the argument for an Elamite connection is probably simply erroneous. Elamite has fragments that resemble

many phyla including Afroasiatic and the apparent cognates probably reflect both trade and migration during a long period and wishful thinking (Blažek 1999). If so, it is likely that Brahui represents a westward migration, not a relic population, especially as the Brahui are pastoral nomads. Kurux and Malto may also be migrant groups, but it seems possible that the centre of gravity of Dravidian was once further north.

Our understanding of Dravidian is strongly related to the Dravidian etymological dictionary of Burrow and Emeneau (1984 and online). However, this is very Tamil-centric and the literature constantly confuses its head entries with proto-Dravidian (e.g. Krishnamurti 2003). Notwithstanding these reservations, Southworth (2005, 2006) has undertaken an analysis of this data in terms of subsistence reconstructions with generally convincing results. Broadly speaking, the earliest phase of Dravidian expansion shows no sign of agriculture but (lexically) reflects animal herding and wild food processing. This is associated with the split of Brahui from the remainder. The next phase, including Kurux and Malto, shows clear signs of agriculture (taro production but not cereals) and herding, while South and Central Dravidian have the full range of agricultural production. Fuller (2003) and Southworth (2006) link this to the aptly named South Neolithic Agricultural Complex (SNAC) dated to around 2300-1800 BC in Central India.

AUSTROASIATIC

Austroasiatic has three significant branches in South Asia, Munḍā, Khasian and Nicobarese. These are discrete populations whose historical origins are quite distinct. Map 1 shows the distribution of Austroasiatic.

Six Nicobarese languages are spoken in the Nicobar islands, an archipelago opposite southern Myanmar (Braine 1970; Das 1977; Radhakrishnan 1981). Nicobarese is most closely related to Monic and Aslian and therefore represents a direct migration from Southeast Asia, while Munḍā and Khasian represent an overland connection. MtDNa work on Nicobar populations has also demonstrated close links with mainland Southeast Asian populations (Prasad *et al.* 2001). However, out understanding of Nicobarese is severely compromised by a lack of descriptions of some of its members as well as a virtual absence of archaeology. The most reasonable assumption is that the early Nicobarese migrations arose from the conjunction of Mon speakers with the 'sea nomads' of the Mergui archipelago (White 1922). Nicobarese agricultural terms show cognacy with the broader Austroasiatic lexicon, suggesting that the original migrants were themselves farmers. Indeed, the main islands have derived savannas of *Imperata cylindrica* grasslands which suggest forest clearance by incoming agricultural populations (Singh 2003: 78).

Khasian languages are thought to be most closely related to Khmuic, a branch which includes the Palaungic languages of northern Burma and the Pakanic languages, a now fragmentary and little-known group in south China. This points to an arc of Austroasiatic which must once have spread from the valley of the Mekong westwards across a number of river valleys. The geographic isolation of different Austroasiatic groupings in this region makes it likely that Tibeto-Burman languages subsequently spread southwards and isolated different populations. Figure 2 shows the internal classification of Austroasiatic according to Diffloth.

The Munḍā languages are spoken primarily in northeast India with outliers encapsulated among Indo-Aryan languages in central India (Bhattacharya 1975; Zide and Anderson 2007). It has usually been assumed that Southeast Asia is the homeland of Austroasiatic and the Munḍā languages represent a subsequent migration. The geography of Munḍā does suggest that it was once more widespread in India and has been pushed back or encapsulated by both Indo-Aryan and Dravidian. Indeed the early literature

Map 1 Distribution of the Austroasiatic languages

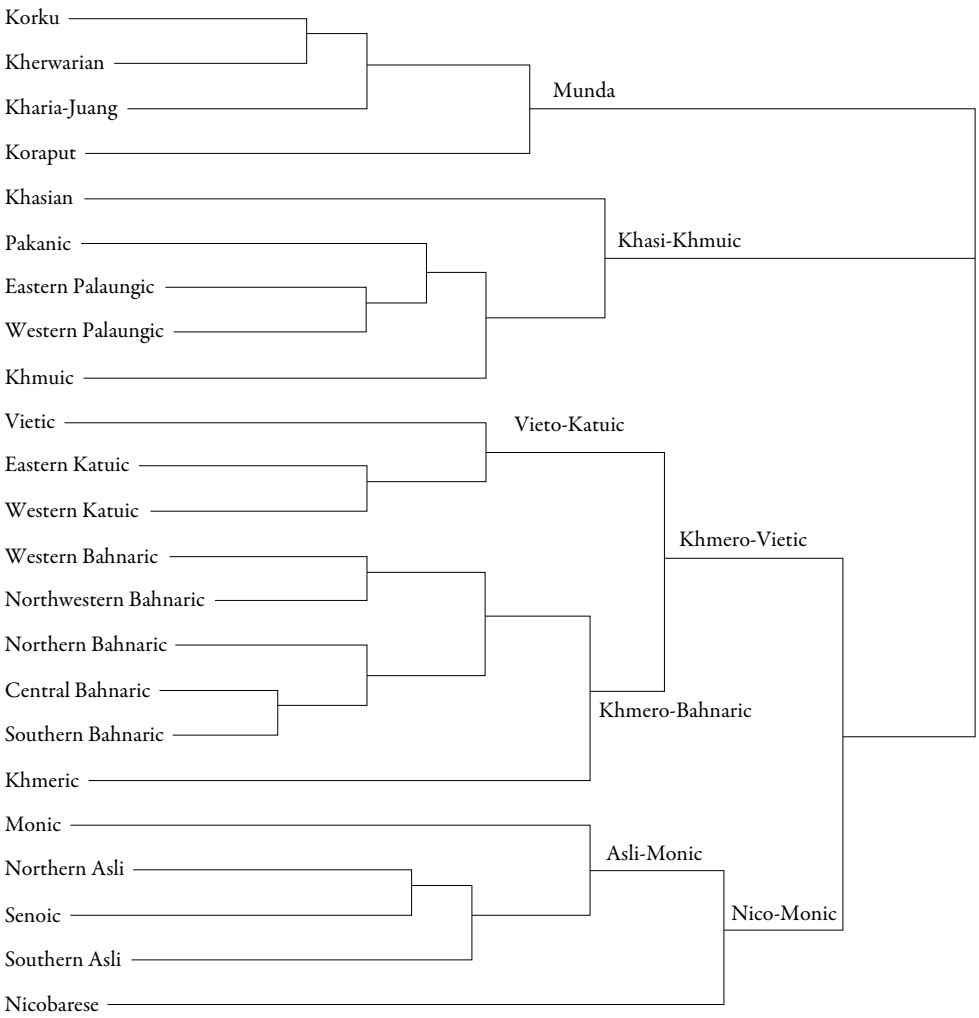

Figure 2 Austroasiatic according to Diffloth (2005)

detected Munḍā influences far to the west of the Indo-Aryan zone even in the Dardic languages of Pakistan (Tikkanen 1988), an idea still countenanced in recent publications (e.g. Zoller 2005). Evidence for this is extremely insubstantial and Munḍā might best be confined to its approximate present region. Munḍā has undergone changes in word order and has other linguistic features that point to long-term bilingualism with non-Austroasiatic languages. Nonetheless, this evidence is susceptible to an opposing interpretation. Donegan and Stampe (2004), for example, argue that the greater internal diversity of Munḍā as opposed to Mon-Khmer imply that it is older and that the direction of spread in Afroasiatic was thus from west to east.

Our understanding of Austroasiatic has been much increased by the publication of Shorto's (2006) comparative dictionary. Proto-Austroasiatic speakers almost certainly already had fully established agriculture. It is possible to reconstruct ox, ?pig, taro, a small millet, numerous terms connected with rice and 'to hoe', all with Munḍā cognates. If so, it seems that Austroasiatic may be 'younger' than the time-scales proposed by Diffloth (2005). The early Neolithic of Southeast Asia, such as that represented at Phung Nguyen (ca. 2500 BP), is associated with rice, domestic animals and forest clearance (Higham 2002). However, our understanding of Austroasiatic is limited by the lack of material on Pakanic and other more remote branches which may represent its earliest phases and such sites may therefore represent a later expansion.

INDO-IRANIAN

Indo-Iranian is the most researched and controversial of the phyla in South Asia, in part due to the rise of nationalist agendas. The Indo-Iranian languages of South Asia are for the most part Indo-Aryan, a category that links together the major languages (including those with a literary tradition) and 100+ 'minor' languages (Nara 1979; Masica 1991; Cardona and Jain 2003). This includes a 'third stream' of the Nuristani languages (in Afghanistan and Pakistan) co-ordinate with Iranian which preserve very archaic features and which remain poorly described. The classification of the Dardic languages remains unresolved, as they may either also be a co-ordinate branch with the others or a primary branch of Indo-Aryan. Figure 3 shows a compromise tree of Indo-Iranian.

The usual model is that Indo-Aryan enters India from the northwest and expands rapidly, bringing with it a host of particular characteristics and assimilating large numbers of Munḍā and Dravidian languages. This does not sit well with nationalist agendas and recent publications have given the in situ hypothesis (that Indo-Aryan is somehow 'indigenous' to India) more credibility than it really deserves. Indeed this has recently been given support by a rather contorted genetic argument (Sahoo *et al.* 2006) which suggests that; 'The distribution of R2, ..., is not consistent with a recent demographic movement from the northwest'. This could also be consistent with the argument that genetics is as much subject to manipulation as any

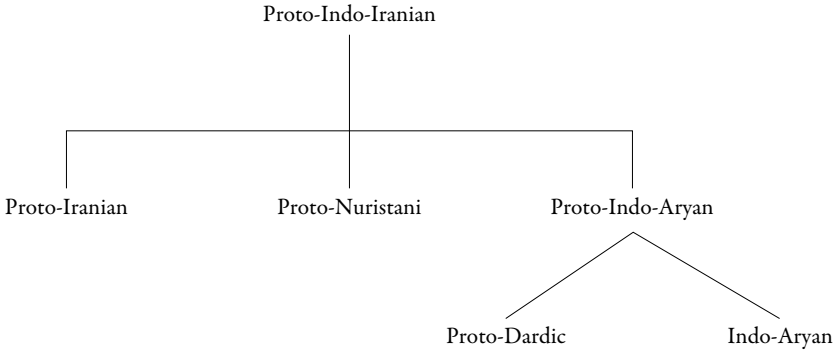

Figure 3 Indo-Iranian 'tree'

other discipline.

The chronology of the Indo-Aryan expansion is still controversial, though it must evidently antedate written attestations. The date of the earliest Vedic scriptures is ca. 1500 BC, so presumably the first appearance of these groups is ca. 2000 BC. Parpola (1988) remains a compelling summary of the archaeological and linguistic evidence. He points out that there is no clear archaeozoological evidence for horses antedating 2000 BC in the Indian archaeological record, and given the centrality of the horse to Indo-Aryan culture, this suggests their presence cannot be significantly older. On the basis of contacts with proto-Finno-Ugric, Parpola places proto-Aryan in south Russia in the middle of the third millennium BC. The first wave of Indo-Aryan migration would then be associated with the spread of Black-and-Red Ware (BRW) which spreads across north India from 2000 BC onwards and which Parpola identifies with the Dāsas of the Ṛgveda. This spread seems also to be strikingly coincident with the appearance of African 'monsoon' crops in the archaeological record. A second wave, characterised by Painted Grey Ware (PGW) overlays BRW from 1100 BC and may be associated with what Grierson called the 'Inner' Indo-Aryan lects, which eventually developed into Hindi. Southworth (2005: 154 ff.) presents an updated interpretation of this hypothesis. The extent to which the expanding Indo-Aryans encountered Dravidian and Muṇḍā speakers is unclear, but it seems certain they assimilated a large number of diverse languages of unknown affiliation spoken by hunting-gathering populations.

The Himalayan range was already occupied by Tibeto-Burman speakers and Indo-Aryan languages must therefore made only a limited impact spreading northwards. Nonetheless, the northern fringe of Indo-Aryan is occupied by a range of diverse languages, many with marked tonal characteristics, suggesting intensive interaction over a long period. Indo-Aryan has loanwords from both Muṇḍā [4] and Dravidian as well as lexemes from presumably now-disappeared language phyla, strongly suggestive of a people moving into a new and unfamiliar environment but interacting with populations who already have agriculture.

Further west, it may be possible to identify Dardic and Nuristani languages with the later Kashmir Neolithic (Fuller 2007). These populations retained non-Muslim religious practices until recently and Parpola (1988: 245) notes that they used ceremonial axes as symbols of rank similar to those on petroglyphs from the upper Indus dating to the 9th century BC. The Nuristani languages have a full suite of 'winter' crops and livestock terms, most of which show cognates with Indo-Aryan proper. The exception is 'millet' (either Setaria or Panicum) which has a diverse range of names strongly suggesting its importance prior to the establishment of the typical Indo-Aryan winter crops. Nuristani languages remain particularly poorly known, with a couple of languages little more than names. Moreover, some Dardic languages, such as Yidgha and Munjani, seem to display particularly striking archaisms and clearly would repay further more detailed study.

SINO-TIBETAN (=TIBETO-BURMAN)

Sino-Tibetan is the phylum with the second largest number of speakers after Indo-European, largely because of the size of the Chinese population. Current estimates put their number at ca. 1.3 billion (Ethnologue 2005). Apart from Burmese and Tibetan, most other languages in the phylum are small and remain little-known, partly because of their inaccessibility. The internal classification of Sino-Tibetan remains highly controversial, as is any external affiliation. The key questions are whether the primary branching is Sinitic (i.e. all Chinese languages) versus the remainder (usually called Tibeto-Burman) or is Sinitic simply integral to existing branches such as

Bodic, etc. as Van Driem (1997) has argued; and what are its links with other phyla such as Austronesian?

Tibeto-Burman studies have been hampered by a failure to publish comparative lexical data and there are thus difficulties in assessing issues such as the early importance of agriculture. The 800-page handbook of Tibeto-Burman published by Matisoff (2003) can only be described as wayward. Among reconstructions it proposes are; 'iron', 'potato', 'banana', 'trousers', 'toast'[!]. The Tibeto-Burman languages (the westernmost of which is Balti in northern Pakistan) have clearly had a significant influence on agricultural vocabulary in the Indo-Aryan languages, as loanwords for 'rice' and some domestic animals indicate.

It is therefore not reasonable at present to reconstruct the history of Tibeto-Burman through either internal genetic classification or comparative lexicon. At present, we can only go by internal diversity and there is no doubt that this is greatest in the Nepal-Bhutan area. The present assumption is that the diverse groups were originally hunter-gatherers making seasonal forays onto the Tibetan Plateau but that 7-6000 BP this became permanent occupation, probably due to the domestication of the yak (Aldenderfer and Yinong 2004). Genetic sampling in Nepal and Bhutan is beginning to make inroads in what has otherwise been a major lacuna.

DAIC (=TAI-KADAI)

South Asia is on the fringe of the Daic-speaking area, which probably originates in south China and may well be a branch of Austronesian. There are some five Tai-speaking groups in northeast India, and oral traditions claim they reached the region in the 13th century (Gogoi 1996). Linguistically, they are a westwards extension of the Shan-speaking peoples of northern Myanmar (Morey 2005). They have a literate culture and individual scripts which relate to the Shan family.

ANDAMANESE

Andamanese languages are confined to the Andaman Islands, west of Myanmar in the Andaman Sea. The Andamanese are physically like negritos, i.e. they resemble the Orang Asli of the Malay peninsula and the Philippines negritos and ultimately Papuans. It has become common currency that the Andamanese are relics of the original coastal expansion out of Africa, and thereby ultimately related to the Vedda, the Papuans and other negrito groups. This has had some recent support from genetics (Forster *et al.* 2001; Endicott *et al.* 2003 [5]) but is still largely unsupported by archaeology (although see Mellars 2006). Some very limited genetic work has been undertaken with the Andamanese. Thangaraj *et al.* (2003) sampled Onge, Jarawa and Great Andamanese as well as museum hair samples but were only able to conclude that the Andamanese were likely to be an ancient Asian mainland population. Although a book about the archaeology of the Andamans has been published (Cooper 2002), in practice it remains unclear when and how the Andaman islands were settled. What few radiocarbon dates exist (Cooper 2002: Table VII:1) are mostly very recent with a small cluster of uncalibrated dates on shell at Chauldari in the 2300-2000 range.

The conversion of Great Andaman to a penal settlement by the British colonial authorities virtually eliminated Great Andamanese and the other languages are severely threatened by settlement from Bengal. Little Andaman (=Onge), Sentinelese and Jarawa are still spoken but Onge, at least, is severely threatened. No data on Sentinelese has ever been recorded and the islanders are officially classified as 'hostile', so classifications of the language are mere speculation. Even the relationship between three partly-documented Andamanese languages is unclear.

Andamanese languages remain poorly documented and statements about their grammar and lexicon difficult to verify. Portman (1898) is the primary early

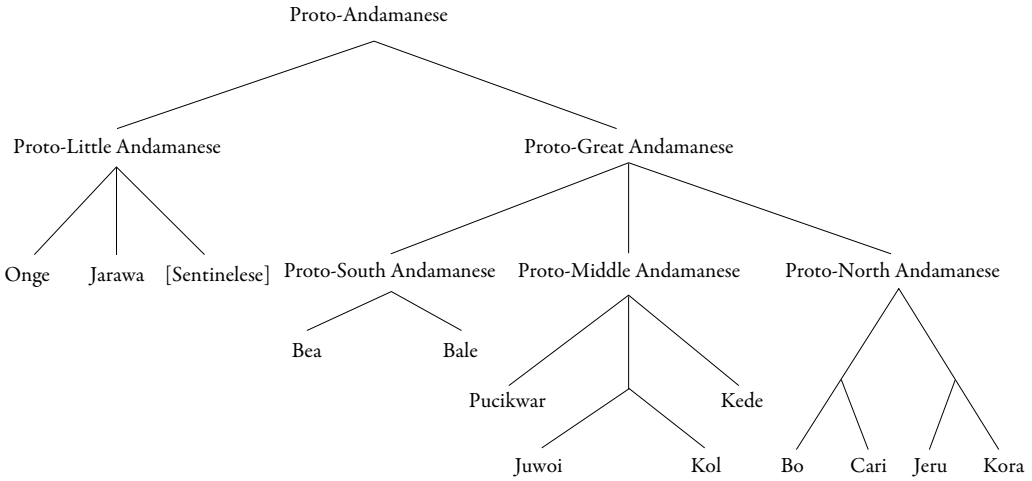

Figure 4 Classification of Andamanese languages

source for Andamanese, and all the available data until 1988 was reviewed by Zide and Pandya (1989) which also contains an exhaustive bibliography. Abbi (2006) represents a partial remedy, providing some basic structural information on three of the four Andamanese languages, but this only serves to deepen the mystery of whether they are in fact related to one another. Some slight resemblances between Andamanese and the 'residual' vocabulary (i.e. non-Austroasiatic) in Aslian have been noted (Blagden 1906; Blench 2006). Andamanese has also been incorporated into the 'Indo-Pacific' model of Greenberg (1971) which despite being reproduced in many reference books and promoted by archaeologists has never garnered significant support from linguists. Blevins (2007) presents new data on Onge and Jarawa, which she claims are related and which she further asserts are a 'Long-lost sister of Austronesian'. The data for Onge and Jarawa do indeed suggest relatively close kinship but the argument for an Austronesian connection is far more tenuous, even given the unlikely prehistoric connection this would suggest. Some of the Jarawa data are quoted from a description of the language in a Ph.D. in progress by Pramod Kumar, so the coming years may see an enhanced understanding of these languages. With reservations, particularly regarding Sentinelese, Figure 4 shows a 'tree' of Andamanese, from Manoharan (1983).

ISOLATES

GENERAL

The four main isolates in South Asia for which significant documentation exists are Burushaski, Kusunda, Nihali and Shom Pen. There is no evidence that these are in anyway related to one another and it is therefore reasonable to think that they are survivors of a period when the linguistic diversity of South Asia was much greater. These languages have been the subject of intensive research by 'long-rangers' but, despite many claims to resolve their affiliation, none have been accepted by a significant body of linguists. Given that most Indo-European specialists think that unknown assimilated languages are a source of aberrant vocabulary in Indo-Aryan it is hardly remarkable that isolates should persist. It may well be that these languages reflect the original hunter-gatherer populations of this region and by considering their agricultural vocabulary we can trawl for indications as to their prehistory.

BURUSHASKI

Burushaski is spoken in the central Hunza valley of northern Pakistan (Backstrom 1992). It is divided into three quite marked dialects, Hunza, Yasin and Nagar. The principal description of the language is Lorimer (1935-38) with additional materials from many other authors (e.g. Tiffou and Pesot 1989). Burushaski has

long been recognised as difficult to classify, and the literature is replete with numerous theories of varying degrees of credibility. Burushaski has been connected with Indo-European, Caucasian, Yeniseian and other phyla (see summary in Van Driem 2001). However, the evidence offered is typically lexical and it is clear that Burushaski has borrowed heavily from a variety of neighbours.

Appendix 1 [6] presents a summary view of crop and livestock vocabulary in Burushaski, with potential etymologies for most words. Burushaski appears to have almost no native crop or livestock vocabulary, but borrows heavily from Dardic and Tibeto-Burman for crops and from Caucasian and Dardic for livestock names. This strongly suggests that the Burushaski were originally hunter-gatherers who adopted agriculture following contact with their neighbours.

KUSUNDA

Kusunda is a language spoken in Nepal by a group of former foragers commonly known as the 'Ban Raja'. It was first reported in the mid-19th century (Hodgson 1848, 1858) but has become known in recent times through the work of Johan Reinhard (Reinhard 1969, 1976; Reinhard and Toba 1970). It was thought to be extinct, but surprisingly some speakers were contacted in 2004 and a grammar and wordlist have now been published (Watters 2005). The language is, however, moribund and high priority should be assigned to developing a more complete lexicon.

There have been numerous claims as to the affiliation of Kusunda, most recently a high-profile (in PNAS) assertion that Kusunda is 'Indo-Pacific' (Whitehouse *et al.* 2004). None of these has met with any scholarly assent and the publication of the Indo-Pacific claim is troubling in terms of non-linguistic journals providing outlets for papers that would not pass normal refereeing processes.

Appendix 2 presents the crop and livestock vocabulary in Watters (2005). In contrast to Burushaski, the existing vocabulary for agriculture seems quite distinctive and only exhibits a few obvious borrowings. Kusunda people appear to be semi-nomadic hunter-gatherers, at least in the recent past, but they may well be a former agricultural group that has reverted to the forest.

NIHALI

The Nihali (=Kolṭu) language is spoken by up to 5,000 people in Maharashtra, Buldana District, Jamod Jalgaon tahsil Subdistrict. Attention was first drawn to this language by the Linguistic Survey of India (Konow 1906) and its exact affinities have long been the subject of speculation. Although the lexicon resembles Korku, a nearby Muṇḍā language with whom the Nihali have a subordinate relationship, there are also extensive loans from the neighbouring Indo-Aryan and Dravidian languages. Speakers use Marathi as a major second language. An overview of the lexicon and its affinities is given in Mundlay (1996) which casts a wide net in seeking the external affinities of Nihali. Carious writers have considered that it is simply aberrant Muṇḍā or some sort of secret language or jargon, but Zide (1996) argues convincingly against these proposals. However, although it is now generally recognised as an isolate, it has been the focus of much theorising, including links with Ainu.

Agricultural vocabulary in Nihali, somewhat strangely, almost all seems to derive from the nearby Indo-Aryan Marathi and not Muṇḍā, as might be expected. Appendix 3 tabulates what can be gleaned from Mundlay (1996) with etymologies given as far as possible. As with Burushaski, the absence of local terms points to a hunter-gatherer group sedentarised under the influence of Indo-Aryan populations.

SHOM PEN

The Shom Pen are a group of some 200 hunter-gatherers inhabiting the centre of Grand Nicobar island. Until recently, the language of the Shom Pen had remained unknown apart from ca. 100 words

recorded by De Roepstorff (1875), the scattered lexical items in Man (1886) and the comparative list in Man (1889). Although most reference books list Shom Pen as part of the Nicobarese languages and Stampe (1966: 393) even stated that Shom Pen is 'possibly extinct', evidence for this is slight. Apart from some numerals and body parts, the Shom Pen words of show no obvious relationship with other Nicobarese languages or other Mon-Khmer languages. The evidence does not immediately suggest that the Shom Pen are Austroasiatic-speakers. Man (1886: 436) says; 'of words in ordinary use there are very few in the Shom Pen dialect which bear any resemblance to the equivalents in the language of the coast people'. Man's Shom Pen d-ata shows that numbers 1-5 are roughly cognate with Nicobarese but that above this they are quite different. Man (1886) also observed that there was substantial linguistic variation between Shom Pen settlements;

> In noting down the words for common objects as spoken by these (*dakan-kat*) people I found that in most instances they differed from the equivalent used by the Shorn Pen of Lafal and Ganges Harbour.

A somewhat difficult to access publication, Chattopadhyay and Mukhopadhyay (2003), makes available a significant body of new data on the Shom Pen language. While not to modern standards of presentation and analysis, it is enough to make a more informed estimate of the affiliation of Shom Pen. The authors consider some of the possibilities and suggest that Shom Pen may be related to Polynesian[!]. Blench (in press) presents a re-analysis of this data and concludes that the evidence points to the status of Shom Pen as a language isolate. He further argues that the marked differences with Man (1889) may point to there being more than one 'Shom Pen' language. Trivedi *et al.* (2006) present some genetic information on the Shom Pen, but without reaching any clear conclusions and certainly without substantiating their conclusion that these are 'descendants of Mesolithic hunter-gatherers'.

CONCLUSIONS

Despite the vast body of research on South Asia, from the point of view of linguistic scholarship, it remains extremely poorly known. This is partly due to restrictions on research, as well as biases that privilege literary languages. Confusions between etymological dictionaries and historical reconstruction underpin false assumptions. The reconstruction of agricultural terminology is beset by poor identifications. More descriptive research and more attention to correct identification of crops, animals and agricultural terminology would improve the potential for correlation with agriculture. This is particularly relevant as it appears that the three most widespread language phyla all began to spread with agriculture already in place.

As for the interdisciplinary reconstruction of South Asian prehistory, the correlations that can at present be hazarded are best described as tentative. The Neolithic archaeology of South Asia is still too poorly known in many areas to make useful links between language and archaeological complexes, even for the most striking expansion, the Indo-Aryan languages. Genetics is also incipient, with exaggerated claims made for restricted datasets. However, none of this is to deny the potential for developing a multi-disciplinary narrative of prehistory; but this will take a renewed impetus in the description of the vast wealth of languages in the South Asian region.

Notes

1) Thanks to Dorian Fuller for both stimulating me to write this paper in the first place and making available a number of unpublished or hard-to-access papers that have been used in composing the text. The paper was first presented at the workshop: Landscape, demography and subsistence in prehistoric India: exploratory workshop on the middle Ganges and the Vindhyas. Leverhulme Centre for Human Evolutionary Studies, University of Cambridge, 2-3 June, 2007. I would like to thank the audience for valuable comments as well as acknowledge additional email comments from Franklin Southworth. George van Driem kindly corrected the transliteration of some of the language data.

2) Southworth gives examples from Dravidian, but this is almost certainly true for other language phyla.

3) Philip Baker has begun work on recovering more of the Wanniya-laeto language and has been able to confirm and extend the materials of earlier researchers. Regrettably, his fieldnotes were destroyed in the 2004 tsunami, so publication may be delayed (Van Driem personal communication).

4) Although the extent of Muṇḍā loanwords may well have been exaggerated. Osada (2006) shows that earlier identifications of supposed Mon-Khmer forms in Sanskrit were in fact the reverse, borrowings into Southeast Asian languages.

5) Although this evidence has been criticised in Cordaux and Stoneking (2003).

6) I am indebted to John Bengtson for an unpublished paper on Burushaski which includes an analysis of links with Caucasian agricultural terminology.

Bibliography

Abbi, Anvita (2006) *Endangered Languages of the Andaman islands*. Lincom Europa, Munich.

Aldenderfer, M. and Zhang Yinong (2004) The prehistory of the Tibetan Plateau to the seventh century A.D.: perspectives and research from China and the West since 1950. *Journal of World Prehistory* 18(1): 1-55.

Backstrom, Peter C. (1992) "Burushaski," in P.C. Backstrom and C.F. Radloff (eds.) *Sociolinguistic survey of Northern Pakistan*, Volume 2. National Institute of Pakistan Studies & SIL, Islamabad. pp. 31-56.

Bengtson, J.D. (2001) Genetic and Cultural Linguistic Links between Burushaski and the Caucasian Languages and Basque. *Paper given at the 3rd Harvard Round Table on Ethnogenesis of South and Central Asia*, Harvard University, May 12-14, 2001.

Bhattacharya, Sudibhushan (1975) *Studies in comparative Munda linguistics*. Indian Institute of Advanced Studies, Simla.

Blagden, C.O. (1906) "Language," in W. W. Skeat and C. O. Blagden, *Pagan races of the Malay Peninsula*, Volume 2. MacMillan, London. pp. 379-775.

Blažek, Václav (1999) "Elam: a bridge between Ancient Near East and Dravidian India?," in R.M. Blench and M. Spriggs (eds.) *Archaeology and Language* Volume IV*: Language change and cultural change*. Routledge, London. pp. 48-87.

Blench, R.M. (2006) Historical stratification of vocabulary in the Aslian languages and potential correlations with archaeology. *Paper presented at the Preparatory meeting for ICAL-3*. EFEO, Siem Reap, 28-29th June 2006.

Blench, R.M. (in press) The language of the Shom Pen: a language isolate in the Nicobar islands. *Mother Tongue* XII.

Blevins, Juliet (2007) "A long-lost sister of Austronesian? Proto-Ongan, mother of Onge and Jarawa of the Andaman islands". *Oceanic Linguistics* 46(1): 154-198.

Braine, Jean C. (1970) *Nicobarese Grammar (Car Dialect)*. Ph.D. Dissertation, University of California, Berkeley

Burrow, T. and Murray B. Emeneau (1984) *A Dravidian etymological dictionary*, second edition. Clarendon Press, Oxford.

Caldwell, R. (1856) *A comparative grammar of the Dravidian or South-Indian family of languages*. Kegan Paul, Trench, Trubner, London.

Cardona, George and Danesh Jain (eds.) (2003) *The Indo-Aryan Languages*. Routledge Curzon Language Family Series, London.

Chattopadhyay, Subhash Chandra and Asok Kumar Mukhopadhyay (2003) *The Language of the Shompen of Great Nicobar : a preliminary appraisal*. Anthropological Survey of India, Kolkata.

Cooper, Zarine (2002) *Archaeology and History -Early Settlements in the Andaman Islands*. Oxford University Press, New Delhi.

Cordaux, R. Weiss, G. Saha, N. Stoneking, M. (2004) "The northeast Indian passageway: a barrier or corridor for human migrations?" *Molecular and Biological Evolution* 21: 1525-1533.

Cordaux, Richard and Mark Stoneking (2003) Letter to the Editor: South Asia, the Andamanese, and the Genetic Evidence for an "Early" Human Dispersal out of Africa. *American Journal of Human Genetics* 72: 1586-1590.

Das, A.R. (1977) *A study of the Nicobarese language*. Anthropological Survey of India, Calcutta.

De Roepstorff (1875) *Vocabulary of dialects spoken in the Nicobar and Andaman islands,* second edition. Calcutta (no publisher).

Diffloth, Gerard (2005) "The contribution of linguistic palaeontology to the homeland of Austro-Asiatic". in Sagart, L. Blench, R.M. and A. Sanchez-Mazas (eds.) *The peopling of East Asia*. Routledge, London. pp. 77-80.

Donegan, P. and D. Stampe (2004) "Rhythm and the synthetic drift of Munda". *The Yearbook of South Asian Languages and Linguistics 2004*. Mouton de Gruyter, Berlin/New York. pp.3-36.

Durie, M. and M. Ross (eds.) (1996) *The comparative method reviewed: regularity and irregularity in language change*. Oxford University Press, New York.

Endicott, P., M.T.P. Gilbert, C. Stringer, C. Lalueza-Fox, E. Willerslev, A.J. Hansen, A. Cooper (2003) The genetic origins of Andaman Islanders. *American Journal of Human Genetics* 72: 178-184.

Ethnologue (2005) URL: http://www.ethnologue.com.

Forster, Peter and Shuichi Matsumura (2005) Did Early Humans Go North or South? *Science* 308 (5724): 965-966.

Fuller, D.Q. (2003) "An agricultural perspective on Dravidian historical linguistics: archaeological crop packages, livestock and Dravidian crop vocabulary," in P. Bellwood and C. Renfrew (eds.) *Examining the farming/ language dispersal hypothesis*. McDonald Institute for Archaeological Research, Cambridge. pp.191–213.

Fuller, D.Q. (2007) "Non-human genetics, agricultural origins and historical linguistics in South Asia," in M.D. Petraglia and B.A. Allchin (eds.) *The evolution and history of human populations in South Asia: inter-disciplinary studies in archaeology, biological anthropology, linguistics and genetics*. Springer Press, Doetinchem. pp. 393-443.

Gogoi, Pushpadhar (1996) *Tai of North-East India*. Chumphra Publishers, Demaji, Assam.

Higham, Charles (2002) *Early cultures of mainland Southeast Asia*. River Books, Bangkok.

Hodgson, Brian H. (1848) On the Chépáng and Kúsúnda tribes of Nepál. *Journal of the Asiatic Society of Bengal* XVII (2): 650-658.

Hodgson, Brian H. (1858) Comparative vocabulary of the languages of the broken Tribes of Népál. *Journal of the Asiatic Society of Bengal* 27: 393-456.

Konow, S. (1906) "Nihali," in G. Grierson (ed.) *Linguistic Survey of India*, Volume 4. Office of the Superintendent of Government Printing, Calcutta. pp.185-189.

Krishnamurti, Bh. (2003) *The Dravidian languages*. Cambridge University Press, Cambridge.

Kumar V, B.T. Langsiteh, S. Biswas, .J.P. Babu, T.N. Rao, K. Thangaraj, A.G. Reddy, L. Singh and B.M. Reddy (2006) Asian and Non-Asian Origins of Mon-Khmer and Mundari Speaking Austro-Asiatic Populations of India. *American Journal of Human Biology* 2006, 18: 461-469.

Kumar, V. and B.M. Reddy (2003) Status of Austro-Asiatic groups in the peopling of India: An exploratory study based on the available prehistoric, Linguistic and Biological evidences. *Journal of Bioscience* 28: 507-522.

Lorimer, D.L.R. (1935-38) *The Burushaski language*. 3 vols. Instituttet for Sammenlignende Kulturforskning, Oslo.

Man, E.H. (1886) A Brief Account of the Nicobar Islanders, with Special Reference to the Inland Tribe of Great Nicobar. *The Journal of the Anthropological Institute of Great Britain and Ireland* 15: 428-451.

Man, E.H. (1889) *A dictionary of the Central Nicobarese language*. W.H. Allen, London.

Manoharan, S. (1983) Subgrouping Andamanese group of languages. *International Journal of Dravidian Linguistics* XII(1): 82-95.

Masica, C. (1991) *The Indo-Aryan languages*. Cambridge University Press, Cambridge.

Matisoff, J.A. (2003). *Handbook of proto-Tibeto-Burman*. University of California Press, Berkeley.

McAlpin, D.W. (1981) *Proto-Elamo-Dravidian: The Evidence and its Implications*. Transactions of American Philosophical Society 71/3, Philadelphia.

Mellars Paul (2006) Going East: New Genetic and Archaeological Perspectives on the Modern Human Colonization of Eurasia. *Science* 313: 796-800.

Morey, Stephen (2005) *The Tai languages of Assam - a grammar and texts*. PL 565. Pacific Linguistics, Canberra.

Mundlay, Asha (1996) The Nihali language. *Mother Tongue* II: 5-47.

Nara, Tsuyoshi (1979) *Avahaṭṭha and comparative vocabulary*

of new Indo-Aryan languages. ILCAA, Tokyo.

Osada, T. (2006) "How many proto-Munḍa words in Sanskrit? - with special reference to agricultural vocabulary," in Toshiki Osada (ed.) *Proceedings of the Pre-symposium of Rihn and 7th ESCA Harvard-Kyoto Roundtable*. Research Institute for Humanity and Nature, Kyoto. pp.151-174.

Parpola, Asko (1988) The coming of the Aryans to Iran and India and the cultural and ethnic identity of the Dāsas. *Studia Orientalia* 64: 195-302.

Pinnow H.-J. (1963) "The position of the Munda languages within the Austroasiatic language family," in H.L. Shorto (ed.) *Linguistic Comparison in Southeast Asia and the Pacific*. SOAS, London. pp. 140-152.

Portman, M.V. (1898) *Notes on the Languages of the South Andaman Group of Tribes*. Superintendent of Government Press, Calcutta.

Prasad, B.V., C.E. Ricker, W.S. Watkins, M.E. Dixon, B.B. Rao, J.M. Naidu, L.B. Jorde and M. Bamshad. (2001) Mitochondrial DNA variation in Nicobarese Islanders. *Human Biology* 5: 715-25.

Radhakrishnan, R. (1981) *The Nancowry word: phonology, affixal morphology and roots of a Nicobarese language*. Linguistic Research Inc., Edmonton, Canada.

Reinhard, Johan G. (1969) Aperçu sur les Kusundâ, peuple chasseur du Népal. *Objets et Mondes, Revue du Musée de l'Homme* IX (1): 89-106.

Reinhard, Johan G. (1976) The Ban Rajas, a vanishing Himalayan tribe. *Contributions to Nepalese Studies - Journal of the Centre for Nepal and Asian Studies, Tribhuvan University, Kîrtipur, Nepal* 4 (1): 1-21.

Reinhard, Johan G. and Tim Toba (=Toba Sueyoshi) (1970) *A Preliminary Linguistic Analysis and Vocabulary of the Kusunda Language*. Summer Institute of Linguistics, Kathmandu. (31-page typescript).

Sahoo, Sanghamitra, Anamika Singh, G. Himabindu, Jheeman Banerjee, T. Sitalaximi, Sonali Gaikwad, R. Trivedi, Phillip Endicott, Toomas Kivisild, Mait Metspalu, Richard Villems and V.K. Kashyap (2006) A prehistory of Indian Y chromosomes: Evaluating demic diffusion scenarios. *Proceedings of the National Academy of Sciences* 103 (4): 843-848.

Sengupta, S., L.A. Zhivotosky, R. King, S.Q. Mehdi, C.A. Edmonds, C.T. Chow, A.A. Lin, M. Mitra, S.K. Sil, A. Ramesh, M.V.U. Rani, C.M. Thakur, L.L. Cavalli-Sforza, P.P. Majumder and P.A. Underhill (2006) Polarity and temporality of high resolution Y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of central Asian pastoralists. *American Journal of Human Genetics* 78: 202-21.

Shorto, H. (2006) *A Mon-Khmer comparative dictionary*. Paul Sidwell, Doug Cooper and Christian Bauer (eds.). PL 579. ANU, Canberra.

Singh, Simron Jit (2003) *In the sea of influence: a world system perspective of [sic] the Nicobar islands*. Lund University, Human Ecology Division, Lund.

Southworth, Franklin C. (2005) *Linguistic archaeology of South Asia*. Routledge Curzon, London/New York.

Southworth, Franklin C. (2006) "Proto-Dravidian agriculture," in Toshiki Osada (ed.) *Proceedings of the Pre-symposium of Rihn and 7th ESCA Harvard-Kyoto Roundtable*. Research Institute for Humanity and Nature, Kyoto. pp.121-150.

Stampe, David L. (1966) Recent Work in Munda Linguistics III. *International Journal of American Linguistics* 32(2): 164-168.

Steever, S.B. (ed.) (1998) *The Dravidian Languages*. Routledge, London.

Thangaraj K., L. Singh, A. Reddy, V. Rao, S. Sehgal, P. Underhill, M. Pierson, I. Frame and E. Hagelberg (2003) Genetic affinities of the Andaman islanders, a vanishing human population. *Current Biology* 13: 86-93.

Thangaraj K, V. Sridhar, T. Kivisild, A.G. Reddy, G. Chaubey, V.K. Singh, S. Kaur, P. Agarawal, A. Rai, J. Gupta, C.B. Mallick, N. Kumar, T.P. Velavan, R. Suganthan, D. Udaykumar, R. Kumar, R. Mishra, A. Khan, C. Annapurna and L. Singh (2005) Different population histories of the Mundari- and Mon-Khmer-speaking Austro-Asiatic tribes inferred from the mtDNA 9-bp deletion/insertion polymorphism in Indian populations. *Human Genetics* 116: 507-517.

Thurgood, Graham and Randy J. LaPolla (eds.) (2003) *The Sino-Tibetan languages*. Routledge Language Family Series. Routledge, London and New York.

Tiffou, É. and J. Pesot (1989) *Contes du Yasin. Introduction au bourouchaski du Yasin avec grammaire et dictionnaire analytique*. Peeters, Leuven.

Tikkanen, Bertil (1988) On Burushaski and other ancient substrata in northwestern South Asia. *Studia Orientalia* 64: 303-325.

Trivedi R, T. Sitalaximi, J. Banerjee, A. Singh, P.K. Sircar and V.K. Kashyap (2006) Molecular insights into the origins of the Shompen, a declining population of the Nicobar archipelago. *Journal of Human Genetics* 51: 217-26.

Turner, R.L. (1966) *A Comparative Dictionary of the Indo-Aryan Languages*. Oxford University Press, London.

Van Driem, George (1997) Sino-Bodic. *Bulletin of the School of Oriental and African Studies* 60(3): 455-488.

Van Driem, George (2001) *Languages of the Himalayas : an Ethnolinguistic Handbook of the Greater Himalayan Region*. 2 Vols. Leiden, Brill.

Watters, David E. (2005) *Notes on Kusunda grammar: a linguistic isolate of Nepal*. National Foundation for the Development of Indigenous Nationalities, Kathmandu, Nepal.

White, W.G. (1922) *The Sea Gypsies of Malaya*. Riverside Press, Edinburgh.

Whitehouse, Paul, Timothy Usher, Merritt Ruhlen, and William S.-Y. Wang (2004) Kusunda: An Indo-Pacific language in Nepal. *PNAS* 101(15): 5692-5695.

Winters, C.A. (2001) Proto-Dravidian agricultural terms. *International journal of Dravidian linguistics* 30(1): 3-28.

Zide, Norman H. (1996) On Nihali. *Mother Tongue* II: 93-100.

Zide, Norman H. and V. Pandya (1989) A bibliographical introduction to Andamanese linguistics. *Journal of the American Oriental Society* 109(4): 639-651.

Zide, Norman H. and Gregory D. S. Anderson (2007) *The Munda Languages*. Routledge Curzon Language Family Series, London.

Zoller, Claus Peter (2005) *A grammar and dictionary of Indus Kohistani*. Volume I: *Dictionary*. Mouton de Gruyter, Berlin/New York.

Zvelebil, K.V. (1997) Dravidian languages. *Encyclopaedia Britannica*. Encyclopaedia Britannica, Chicago. pp.697-700.

Appendix 1: Burushaski agricultural terminology

The core list comes from a paper by John Bengtson (2001) which was prepared with the motive of demonstrating the links between Burushaski and Caucasian. I have added other Burushaski terms from Backstrom (1992) and Lorimer (1935-38) as well as adapting materials from other volumes in this series to add to the etymologies. Standard online dictionaries were used for terms in major languages. I do not endorse all Bengtson's connections but I have left most of them in place for discussion, while adding other possible entries to the commentary.

Burushaski	Gloss	Etymological commentary
Livestock		
aćás (H,N,Y)	sheep, goat [1] = *Kleinvieh*, small cattle	cf. Shina *aai* 'goat', Cauc: Adyge *āča* 'he-goat', Dargwa (Akushi) *ʕeža* ~ (Chirag) *ʕač:a* 'goat', etc.< PNC **ʔējʒwē* (NCED 245)
bɛpuy	yak	cf. Shina *bɛpo*, Kohistani *bhép*^{b},
bʌskarẹ̣t	ram	?
buʾa	cow	cf. Balti, Tibetan *ba* ( )
buć (H,N)	(ungelt) male goat, 2 or 3 years old'	cf. Wakhi buč, but Indo-European. cf. Sanskrit *bōkka* and Fr. *bouc*, E. *buck*. Also Cauc [2]: Lak *buχca* (< **buc-χaʔ*) 'he-goat (1 year old)', Rutul *bacʾi* 'small sheep', Khinalug *bacʾiz* 'kid', etc. < PEC **b[a]cʾV* (NCED287). Possibly also Niger-Congo, cf. Common Bantu -búdì
buš	cat	< IA languages
bʌyum	mare	cf. Shina *bʌm*

Burushaski	Gloss	Etymological commentary
čigír (Y) ~ *č higír* (N) ~ *čhiír* (H)	(she-)goat	~ Cauc: Karata *č':ik'er* 'kid', Lak *č'uku* 'goat', etc. < PNC **ʒikV̌/ *kiʒV̌* (NCED 1094)~ Basque *zikiro* ~ *zikhiro* 'castrated goat'
čhindár (H,N) ~ *čuldár* (Y)	bull	~ Cauc: Chamalal, Bagwali *zin*, Tindi, Karata *zini* 'cow', etc. < Proto-Avar- Andian **zin-HV* (NCED 262-263) ~ Basque *zezen* 'bull' (Yasin form influenced by *čulá*? See next entry.)
čhulá (H,N) ~ *čulá* (Y)	male breeding stock': (H) drake, (N,Y) 'buck goat'	cf. Sau *čɔli* 'goat', Cauc: Andi *č'ora* 'heifer'
čhərda	stallion	cf. Shina *čhərda*
du (H,N,Y)	'kid, young goat up to one year'	cf. Cauc: Chechen *tō* 'ram', Lak *t:a* 'sheep, ewe', Kabardian *t'ə* 'ram', etc. < PNC **dwănʔV* (NCED 405)
ḍágar (N)	ram	~ Cauc: Avar *deʕén* 'he-goat', Hinukh *t'eq^{w}i* 'kid (about 1 year old)', etc. < PEC **dVrq'wV* (NCED 403)
élgit (N) ~ *hálkit* (Y)	she-goat, over 1 year old, which has not given birth	~ Cauc: Agul, Tsakhur *urg* 'lamb (less than a year old)', Chamalal *barg^{w}* 'a spring-time lamb', etc. < PEC **ʔwilgi* (NCED 232)
ɣʌla	flock	< Farsi
haġúr ~ *haġór*	horse	~ Cauc: Kabardian *x^{w}āra* 'thoroughbred horse', Lezgi *χ^{w}ar* 'mare', etc. < PNC **farnē* (NCED 425) Also Turkish *aiġır* 'stallion'.
hišmahiiš (H)	buffalo	cf. Wakhi *išmayvš*
hər (in compounds)	bull	?
huk	dog	?
hʌlden	full-grown goat	?
huo	sheep	?
huyés (H,N,Y)	small ruminants	?
jʌkun	donkey	cf. Shina *jʌkun*
qarqaamuš	chicken	cf. Shina *karkamoš*, Wakhi *k^{h}ɛrk*
thugár (H,N)	buck goat	cf. Wakhi *t^{h}uɣ* 'goat', also Cauc: Karata *t'uka* 'he-goat', Bezhta *t'iga* 'he-goat', etc. < PNC **t'ugV* (NCED 1003)
tilíaṅ (H,N) ~ *tilíha ṅ* ~ *telého ṅ* (Y)	saddle (n.)	Cauc: Avar *ƛ':ilí* [*ɬ':ilí*], Lak *k'ili*, etc. 'saddle' < PEC **ƛ'wiɬē* 'saddle' (NCED 783)
xuk	pig	< Farsi
Crops and agriculture		
aalu	potato	< IA languages
aam	mango	< IA languages
balt	apple	cf. ? Wg. *palā* apple
baẏ (H,N: double plural baçéy ~ *báẏiṅ*) ~ *ba* (Y) also *bʌẏ*	millet (Panicum miliaceum)	~ Cauc: Chechen *borc* 'millet', Karata *boča* 'millet', etc. < PNC **bŏlćwĭ* (NCED 309)
ba'logʌn	tomato	
beŋgʌn, *pʌṭigan*	eggplant, brinjal	cf. Hindi *baiṅgana* (बैंगन)
bəru	buckwheat	cf. Shina *bərao* perh. also Sanskrit *phapphara*
bičil	pomegranate	?
birʌnč	mulberry	cf. Shina *maroč*
boqpʌ (H,N)	garlic	< Shina *bokpa,*
bukʌk	beans	< Shina *bukʌk*
bupuš	pumpkin	
brʌs, *briu*	rice	< Balti, Tibetan *'bras* ( )
bʌdʌm	almond	< Farsi
buwər	water-melon	cf. Shina *buwər*
ćha (H,N) ~ *ća*	millet (Setaria italica)	? < Indo-Aryan cf. Gujarati *kāŋg k.o.* grain, Marathi *kāg Panicum italicum*. ? Caucasian: Bezhta *č'e* 'a species of barley', Andi *č'or* 'rye', etc. < PEC **č![e]ɬlV* (NCED 384)
čoṭʌl	rhubarb	cf. Shina *čōṭʌl*

Burushaski	Gloss	Etymological commentary
daltán- (N) (< **r-aƛa-n-*)	'to thresh (millet, buckwheat)'	~ Cauc: Ingush ard-, Batsbi arl- 'to thresh', Tindi rali 'grain ready for threshing', etc. < PEC *-V̄rλV 'to thresh', *r-ɛ̆λe 'grain ready for threshing' (NCED 1031)
darċ	threshing floor, grain ready for threshing	~ Cauc: Dargwa daraz 'threshing floor', Lak t:arac'a-lu id., Tabasaran rac: id., etc. < PEC **ẖrənʒū* (NCED 503)§ Comparison by Bouda (1954, p. 228, no. 4: Burushaski + Lak).
ḍoŋhər	mustard	cf. Shina *doŋhʌr*
ġašu	onion	< Shina *kašu*,
ġ̇ərk	peas	cf. Kashmiri *kala* pea (Pisum satvum)
gobi (H, N, Y)	cabbage	< IA languages, e.g. Hindi *gōbhī* (गोभी)
grinč (Y)	rice	cf. Khowar *grinč*, Wakhi *gʌrɛnč*
gur (H,N,Y)	wheat	Tibetan *gro* ( ) 'wheat' also Cauc: Tindi *q':eru*, Archi *qoqol*, etc. 'wheat' < PEC **ɢōlʔe* (NCED 462) ~ Basque *gari* 'wheat' (combinatory form *gal-*)
gʌškur	cherry	?
ɣon	musk-melon	?
haḷiċi (N)	turmeric	< IA languages, e.g. Shina *haliči*, Hindi *haladī* (हलदी)
harṣ (H,N) ~ *harṣ*~ *haṣç* (Y)	plough	~ Cauc: Akhwakh ʕerc:e 'wooden plow', Lak qa-*ras* id., etc. < PNC **Hrājcū* (NCED 601)
həri	barley	?
jotu	chicken	cf. Shina *joṭo*
ju	apricot	?
jʌṭor	quince	cf. Shina *čʌṭor*
limbu	lemon	< IA languages
maručo	chili	cf. Hindi *mirca* (मिर्च) but perhaps via Wakhi *mʌrč*
mumphʌli	groundnut	cf. Hindi *mūm̃gaphalī* (मूँगफली)
pfak	fig	cf. Sh. *phāg* but widespread in Indo-European and ultimately E. 'fig'
phɛso	pear	?
šinaba'logʌn	eggplant, brinjal	name of 'tomato' + qualifier (q.v.). However, a similar formation occurs in Shina, and the qualifier *kiṇo* means 'black' suggesting this expression is borrowed from Shina
siŋur (H)	turmeric	?
tili	walnut	?
ṭuru	pumpkin	? cf. Shina *turu* 'small bowl'
wʌžnu (Y)	garlic	cf. Kalash wɛšnu,
zɛsčʌwa (Y)	turmeric	cf. Khowar-Khalash *zɛhčawa*,

This analysis suggests that there is no proof Burushaski is not genetically related to any of the phyla in which cognates have been detected, but rather that the original Burushaski were not farmers or even herders but hunter-gatherers, who built up their agriculture by borrowing from a wide variety of neighbouring peoples.

Appendix 2:

Crops in Kusunda

Kusunda	English	Etymology
əmbyaq	mango	< Nepali *ambak~amba* guava; cf. also *ãp* mango
əraq / əraχ	garlic	
abəq / əboχ	greens, vegetable	
begəi	ginger	
begən	chilli, pepper	
byagorok	radish	
dzəpak	k.o. yam	
gisəkəla	oats	identification of cultigen uncertain in source

Kusunda	English	Etymology
goləŋdəi	soya bean	
ghəsa~gəsa	tobacco	
ipən	corn, maize	
kəpaŋ	turmeric, besar	
khaidzi	food, cooked rice	
laī / lāɲe, ləŋkan	cucumber	
motsa	banana, plantain	cf. Arabic mōz
nimbu	lemon	Terai Nepali *nimbu* निंबु
pəidzəbo	black gram	equivalent to Nepali *mās* मास
pyadz	onion	Nepali *pyāj* प्याज
pheladəŋ	lentil	equivalent to Nepali *gagat* गगत
phelāde	beaten rice	
rəŋgunda / rəmkuna	pumpkin	
rāko, raŋkwa	millet	
rambenda	tomato	<Nepali *golbhēḍā* गोलभेंडा
ran	millet	
sen	paddy	
simi	beans	<Nepali *simī* सिमी
yebu	yam	

Domestic animals

Kusunda	English	Etymology
agəi	dog	
aidzi / əidzi	goat	
gya numba	ox	
hi / he	pig, wild boar	
məhi, məih	buffalo	cf. Tibeto-Burman *mes; cf. Skt XXX.
numba	cow	
tap	chicken	

Others

Kusunda	English	Etymology
ghue ə-go	to plough	
sisen / sisin	paddy, rice field	

Appendix 3: Nihali agricultural vocabulary

Nihali	Gloss	Etymology	Comment
āndij	root like sweet potato		
āṭa	food	cf. Hindi *āṭā* आटा 'wheat flour'	
bābā	paddy	< Korku *baba*	
bābarā	edible root		
baḍágo	guava		New World
baḍḍi	bull		
bakārā	he-goat	< Marathi *bakarā* बकरा	
bardo	sickle	?	
baru	mulberry		
bātuko	mango	<Marathi *bitki* 'mango-stone'	
bhaji	vegetables	<Marathi (and Common IA) *bhaji*	
bhedarā	potato		New World
bhendye	okra	< Marathi *bhende*	
bher	to harrow		
bolor	edible root		
cāpir	edible root		
chabbāl	iron hoe		
chunḍu	bean		
cicca	tamarind	? < Marathi but widespread in Dravidian and Muṇḍā and borrowed into Sanskrit	
co(g)gom	pig		
dāngārā	cucumber	< Marathi *dangar* 'gourd'	
dāwrā	harrow	< Marathi *davrā* 'plough'	
dhor	cattle	< Marathi *dhor*	
dole	cardamom	< Marathi *veldode*	
dotkā	k.o. gourd	< Marathi *doḍka*	
ḍotako	edible root		
engan	eggplant	< Hindi *baiṅgana* बैंगन	

Nihali	Gloss	Etymology	Comment
gadri	donkey	< Marathi *gāḍhava* गाढव	
gājre	carrot	< Marathi *gājar* गाजर	
gele	maize	< Korku *gele* 'ear of maize'	New World
gohū	wheat	< Marathi *gahu*	
gorha	male calf	< Marathi guḍghā गुड़घा	
hardo	turmeric		
hellā	male buffalo	< Marathi *hela*	
ilāyci	cardamom	< Hindi *ilāyacī* इलायची	
irā	sickle	<Korku *(h)ir*	
jiryāngā	tomato		New World
kānde	onion		
kapcho	cotton	<Marathi *kāpūs* कपूस	
karelā	bitter gourd	< Hindi *karelā* करेला	
karjo	cashew	< Marathi *kājū* काजू	New World
kelli	female calf	< Korku *kelli*	
khuḍe	gourd		
kombā	cock	< Korku *kōba*	
koplyā	hoe		
longo	clove	cf. Marathi *lavaṅg* लवंग	
lusun	garlic	< Marathi *lasūṇ* लसूण	
malkānbijo	bean		
māre	colt		
masur	red lentil	< Marathi *masūr* मसूर	
methi	fenugreek	<Marathi *methī* मेथी	
mircān	chilli	< Marathi *mircī* मिरची	
mongo	green gram	cf. Marathi *mūng*	
nāy	dog	< Tamil *nāy* நாய்	
nimbu	lime	< Hindi *nībū* नीबू	
odow	female buffalo		
oró	millet	? but cf. Tamil *uruppam* 'millet flour' உரப்பம்	
phellyā	groundnut	cf. Hindi *mūṅgaphalī* मूँगफली	New World
photre	chilli		New World
seri	goat	< Marathi *śeḷī* शेळी	
sitā	dog		
sitāphal	custard apple	< Hindi *sītāphala* सीताफल	
sonu	black-eyed beans		
tāndur	cooked rice	? < Hindi 'clay oven' *tandūr* तंदूर	
turi	gram		

Source of Nihali data is Mundlay (1996)

Notes to Appendix

1) The semantic variation 'sheep' ~ 'goat', which recurs in several of the comparisons below. Cf. huyés, below.

2) 'Young goat' and 'young kid' are polysemous within East Caucasian.