



Origin and Development of Language in South Asia: Phylogeny Versus Epigenetics?

Citation

Witzel, Michael E. J. Origin and development of language in South Asia: Phylogeny versus epigenetics? Paper presented at Darwin and Evolution, mid-year meeting of the Indian Academy of Sciences, Hyderabad, India, July 3, 2009.

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:8554510>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

2

Origin and development of language in South Asia: Phylogeny versus epigenetics?

MICHEL WITZEL

Department of Sanskrit and Indian Studies,
Harvard University, 1 Bow Street,
Cambridge MA 02138, USA
Email: witzel@fas.harvard.edu

SUMMARY

This presentation begins with a brief overview of opinions on the origin of human language and the controversial question of Neanderthal speech. Moving from the language of the ‘African Eve’ to the specific ones of the subcontinent, a brief overview is given of the prehistoric and current South Asian language families as well as their development over the past c. 5000 years.

The equivalents of phylogeny and epigenetics in linguistics are then dealt with, that is, the successful Darwinian-style phylogenetic reconstruction of language families (as ‘trees’), which is interfered with by the separate wave-like spread of certain features across linguistic boundaries, even across language families. A combination of both features leads to the emergence of the current South Asian linguistic area (sprachbund). This development has made the structure of Indo-Aryan, Dravidian or Munda similar to each other but it could not eliminate most of

their individual characteristics.

I. INTRODUCTION

We should increasingly look beyond the narrow confines of our respective disciplines, as discoveries in one field of science have important bearings on other fields --- and this jubilee meeting devoted to Charles Darwin's memory is a welcome opportunity to do so. As announced, this 'Symposium [is] to analyse the concept of evolution or change as understood in various disciplines, and if possible with an Indian context in mind'. I will restrict myself to talking about the linguistic and cultural aspects of South Asian humanity, but this automatically involves taking a look into other fields such as archaeology and genetics, not to speak of ancient texts and belief systems.

Just as in biology, changes in language occur in very small steps: the steps are 'random' (i.e. occur in no specified direction); they can be passed on from one generation to the next, almost always from parent to offspring. And finally, though not immediately visible to the casual observer, if their outcome happens to be of advantage to the individual in whom they occur, they spread in the population for reasons of prestige, etc. (which, ultimately due to societal pressures, also creates a reproductive advantage). Language development agrees with what Theodosius Dobzhansky wrote in 1973: 'nothing in biology makes sense except in the light of evolution'. Language is, after all, an epiphenomenon of the behaviour of the human and some other kinds of apes, some of whom now appear to have some primitive forms of vocal signs that amount to rudimentary speech with syntax (see below), something that has so far been thought to be restricted to humans.

2. ORIGINS?

The origin of language has been discussed at least since the Egyptian pharaoh Psammeticus (664–610 BCE) who, according to Herodotus II 2, isolated two children with a shepherd from the time of birth and concluded, from the first word (bekos 'bread') they spoke to each other, that the original language of all humans

was Phrygian, an Anatolian language.¹ This is of course as wrong as the idea of some that all languages of India – or even of the whole world – are derivatives of Sanskrit. A 16th century Dutchman has claimed that honor for his mother tongue, Dutch. We have older speculations of that kind in the oldest Indian text, the Rigveda.²

We do not know about ultimate origins of language, but speculation has been rife, so that the Linguistic Society of Paris forbade discussion on this point in its meetings in the mid-nineteenth century. Nonetheless, people have gone on discussing this point, and there are strong opinions, some allied to the American linguist Noam Chomsky,³ some not. Chomsky maintains that language faculty is inborn in newborn babies, including even a General Syntax – that is the way words in a sentence are ordered. They would not have to learn such features from their parents. Others such as Philip Lieberman (2006, 2008) maintain that language developed over time, like any other human trait, and that the development of speech in babies follows a gradual pattern, just as they develop other skills. In short, there is no ‘inborn’ universal grammar.

Apart from the fact that the 6000 or so remaining human languages have developed a large variety of possibilities of arranging individual words or their close collocations/synthetic units, the case has not been made that there is an underlying universal grammar. Many diverse types of grammar exist, for example, some without the familiar word classes such as nouns (Hopi) or with tense-based adjectives (Japanese), all which is obvious if one compares Indo-European with Chinese, Eskimo (Inuit), Hopi or Polynesian. An underlying universal grammar is also not found in pidgins and creoles that have evolved more recently: they are based on one dominant language at the time of formation, such as Portuguese, English or others such as Amerindian, in the Chinook jargon⁴ of northwestern America.

1 Obviously, *bek-* is close to what the Greeks and many other people think sheep ‘say’: *bé-bé* (or *mé-mé*), so there is no wonder that these children said *bé-*

2 RV 1.164.45 *catvāri vāk parimitā padāni* (also in Patañjali’s *Mahābhāṣya*)

3 Bickerton 1990

4 It is based on northwestern American Indian words and became a contact language in a very varied geographical and linguistic area, with participation of

Lieberman and others⁵ trace the origin of language to the development of certain neural bases in the brain, not in one region (Broca's area) as was thought in the nineteenth century, rather, in 'circuits' connecting different parts of the brain. This is now generally agreed to by neurologists. These circuits govern motor control of the body, but also cognitive processes in humans, including input from hearing and producing language. Lieberman thus investigates the development of speech and motor control in human ancestors and draws a clear distinction between even the great apes and humans.⁶

However, humans and the great apes differ in the development of speech faculty. Apes generally do not proceed beyond the abilities of a 2–3-year-old infant when using sign language. They can acquire a passive vocabulary of several hundred (or even more) words (signs), without an obvious syntax, the word order within a given sentence that is universally found in all human speech. Conversely, most recent research has indicated that Campbell monkeys may indeed possess syntax in their vocal exchanges in the wild, combining six basis calls in various ways, as the occasion requires.⁷

The question of the original development of speech is one of selection advantage, see Darwin (1859): 'structures that were

a dozen languages and families.

5 Lieberman & McCarthy: Summer 2007:

<http://penn.museum/documents/publications/expedition/PDFs/49-2/Lieberman.pdf>; longer paper in Science Direct: CORTEX 44, 2008, 218-226 .

6 This is also seen in the FXP2 gene = that is, however, not a 'language gene', see later on.

7 From the Summary:

<http://www.pnas.org/content/106/51/22026.abstract?etoc> 'Adult males produced six different loud call types, which they combined into various sequences in highly context-specific ways. Callers followed a number of principles when concatenating sequences, such as nonrandom transition probabilities of call types, addition of specific calls into an existing sequence to form a different one, or recombination of two sequences to form a third one. These primates have overcome some of the constraints of limited vocal control by combinatorial organization. The Campbell's monkey call system may be the most complex example of "proto-syntax" in animal communication known to date' - We need to study whale and dolphin signals more closely than has been done so far. (see: however, http://www.dichotomistic.com/mind_readings_dolphin_speech.html).

initially adapted to control one function take on new “tasks” in the course of evolution’. What then about our closest relatives, the Neanderthals?

3. NEANDERTHALS

In Europe, these early humans are known as Neanderthals or as *Homo neanderthalensis* that developed from *Homo heidelbergensis* and existed by 370 000 BCE, while archaic *Homo Sapiens sapiens* (AMH) first appeared in African fossil remains some 160 000 years ago and in Europe only at c. 40 000 BCE.

In recent years, Neanderthal DNA has been extracted; it shows comparatively little genetic variation with *Homo sapiens*: both are 99.5% identical, though the remaining 0.5% may still result in large differences, as is clear when comparing humans with their close, 98% identical Chimpanzee relatives (Marks 2002). However, Neanderthals are not identical with modern humans, and Neanderthal DNA does not survive in ours, if interbreeding had indeed taken place.⁸ If so, their descendants must have died out.

For the present purpose, the Neanderthal faculty of speech, if any, would be of some interest.⁹ While modern humans and Neanderthals already share 99.5% of their genes, Neanderthal anatomy suggests to my colleague Ofer Bar-Yosef¹⁰ that Neanderthals could speak: a skeleton excavated in an Israeli excavation at Kebara II in 1983 (60 kya),¹¹ has a hyoid bone that is necessary for human-like speech. This bone (with attached

8 However, survival of Neanderthal genes in us has recently been claimed by Pääbö et al., based on old Neanderthal genetic materials. For lack of genetic evidence see Schmitz 2003; s. Schrenk & Müller 2005: 110; cf. Culotta 2007; Noonan 2006.

9 Lieberman 2006, 2007, Schrenk and Müller 2005. For an update on Neanderthal speech, see: http://www.newscientist.com/article/dn13672-neanderthals-speak-out-after-30000-years.html?DCMP=ILC-hmts&nsref=spectr10_head_Neanderthal%20talk.

10 Harvard Magazine Sept./Oct. 2001, <http://www.harvard-magazine.com/on-line/09016.html>.

11 See preceding note, and cf. Schrenk and Müller 2005: 81.

muscles) allows the tongue to modify the space in the throat that is needed for proper articulation.

The question of Neanderthal speech, however, rests on the specifics of the Neanderthal larynx, which has not survived in fossils.¹² Available skeletal remains indicate that it apparently was not in the right position to produce our type of fully vocalized human speech. This would agree with Lieberman's position,¹³ who points out that Neanderthal neck lengths are too short for a 1:1 relationship of SVTh::SVTv (supralaryngeal vocal tract – horizontal::vertical). But, the position of the larynx he posits for Neanderthals seems very close to AMH to me, as it did to Boë et al. (2001) who have pointed out that 'the potential Neanderthal vowel space was as large as that of modern humans'. Also note that a hyoid Neanderthal bone has been found in an Israeli excavation in 1984.¹⁴ The hyoid bone is necessary to keep the tongue in a position that allows production of human-like speech.¹⁵

Another factor that has been brought up repeatedly in the discussion is the so-called, or rather mis-called 'language gene'" *FOXP2*.¹⁶ But, this gene is linked to many functions and only

12 For an update on Neanderthal speech, see: http://www.newscientist.com/article/dn13672-neanderthals-speak-out-after-30000-years.html?DCMP=ILC-hmts&nsref=spectrt10_head_Neanderthal%20talk; Lieberman 2006; Schrenk & Müller 2005, Junker 2006: 99 sqq.

13 Repeating what he and Crelin had already stated in 1971. See Lieberman & McCarthy: Summer 2007: <http://penn.museum/documents/publications/expedition/PDFs/49-2/Lieberman.pdf>, and in *Cortex* 44, 2008, 218-226 (in answer to Daniel Bub's review of his 2006 book "Toward an evolutionary Biology of Language." Bub is a Chomskian, who believes in pre-recorded syntax in babies.)

14 Arensburg et al. 1989: 'dating from about 60 000 years BP. The bone is almost identical in size and shape to the hyoid of present-day populations, suggesting that there has been little or no change in the visceral skeleton ... during the past 60 000 years of human evolution. We conclude that the morphological basis for human speech capability appears to have been fully developed during the Middle Palaeolithic.'

15 Cf, Callaway 2008.

16 Enard W. et al. 2009, referenced in Science Daily: <http://www.sciencedaily.com/releases/2009/06/090624093315.htm>.

marginally to speech development.¹⁷ Then, there is the recent¹⁸ discovery of a primate variant of the *FOXP2* gene in ancient Neanderthal DNA.¹⁹ Both this and our own *FOXP2* variant, then only points to general speech ability.

Indeed, Maria Agnes Solymosi et al.²⁰ point out that (a) *FOXP2* is preserved among mammals: chimp, gorilla, rhesus have identical *FOXP2*, and one amino difference from the mouse, but two from human protein. Orang Utan has two differences from mouse, and three from humans. (b) Only humans have two different amino acid substitutions.²¹ This would indicate that the single human substitution alone is insufficient for acquiring speech and language, between c. 200 kya–100 kya.

However, Neanderthal genomics now indicate that they had our *FOXP2* variant. This change had occurred already at 300–400 kya ago, with the common ancestor of Neanderthals and ANM humans. Could Neanderthals then speak like we do?

Indeed, there was apparent trade exchange between *Homo sapiens sap.*, who entered Europe from the Near East about 40 kya,

17 It is not ‘*the language gene*’ but one of many involved in speech, and it has been present in mammals for 70 million years, such as in mice or bats, just as in the great apes, all of which do not use spoken language. *FOXP2*’s relevance for speech is now questioned even by its co-discoverer, Simon Fisher (2006) who denies a ‘language gene’: ‘Genes do not specify behaviours or cognitive processes; they make regulatory factors, signaling molecules, receptors, enzymes, and so on... much of the data on *FOXP2* from molecular and developmental biology confounds any expectations that one might have for a hypothetical “language gene”.’ Alec MacAndrew (2002) sums up that the development of language did not rely just on a single mutation in *FOXP2* and that many other changes were involved, such as anatomical ones of the supralaryngeal tract. He stresses that all of this did not occur over just 100 000 years. Further, it ‘involved many more genes that influence both cognitive and motor skills ... Ultimately, we will find great insight from further unraveling the evolutionary roots of human speech – in contrast to Noam Chomsky’s lack of interest in this subject.’

18 For a recent discussion of the *FOXP2* gene, see Solymosi et al. 2007; for the emergence of anatomically modern human behaviour in S. Asia, see H. James in Petraglia and Allchin 2007: 204 sqq.

19 Krause 2006, Trinkhaus 2007

20 Solymosi et al. 2007

21 Threonine to asparagine at position 303, and asp. to serine at 325. Other mammals and one bird tested have a *FOXP2* variant (threonine asparagine subst.) with exc. of carnivora (dog to sea lion) at 303, 325.

and Neanderthals. It is seen, for example, at Vindija in Croatia.²² Such exchanges may require rudimentary speech but one can also think of silent exchanges, as has been typical for initial human contact between two mutually alien groups, such as is still seen in the Andamans (Sentinel Islanders).

Some remnants of Neanderthal rituals found in graves²³ also point at symbolic thought that is necessary for speech. Neanderthal burials in Shanidar (N. Iraq) indicate a clear perception of death and the intention to preserve the ‘life force’ of the deceased by putting ochre colour (though not flowers!) on the body.²⁴ These and a number of widely dispersed other finds point to some Neanderthal religious or mythical concepts of an afterlife.

Leaving aside the Neanderthals, Lieberman²⁵ maintains that even anatomically modern humans did not possess full vocalized speech until c. 50 kya. The owner of the ANM Shkul (Israel) skeleton of 90 kya was, in his opinion, not fully able to speak as we do. I find this and some of his statements²⁶ hard to swallow, suppressed larynx or not. The same applies to his statement, following the fact that a baby’s tongue gradually descends into the pharynx, continuing until they are 6–8 years old: ‘At this point we are able to produce the quantal vowels [a] [i] [u]’ and ‘without them speech would still be possible but less effective’.²⁷ However, babies can clearly pronounce [a, i, u] even before they are 2 years

22 Schrenk and Müller 2005: 112. However, other forms of contact, such as in silent trading, or exchanges based on very limited faculty of speech, may also have taken place.

23 Schrenk & Müller 2005: 96 sqq, 108.

24 The insertion of flowers into Neanderthal graves is a modern myth. The famous Shanidar grave in N. Iraq was contaminated: the pollen of flowers found there has been brought down to these levels by rats. (Schrenk & Müller, 2005: 80). On the other hand, there certainly was Neanderthal ritual, such as the bear cult and their death ceremonies bear out.

25 Lieberman and Robert McCarthy’s’ Articles in Museum U. Penn., cf. http://www.newscientist.com/article/dn13672-neanderthals-speak-out-after-30000-years.html?DCMP=ILC-hmts&nsref=specrt10_head_Neanderthal%20talk; in detail Lieberman 2006.

26 Cf. Bub 2008: 20.

27 Lieberman and McCarthy 2007/2008: 18/20.

old.²⁸

In sum, we are left with the fact that both Neanderthals and early AMHs had the physical capability to produce human-like speech.

AFRICAN EVE

As all humans on this planet are closely linked genetically and in fact go back to a single woman in Africa, the ‘African Eve’ (and similarly to a male ancestor), we may try to make some statement about their speech. Over the past two decades, population genetics has increasingly indicated that the story of the spread of humans is one of migration out of Eastern Africa. Second, that several bottlenecks were faced by our early ancestors: the original African exodus by some 2–10 000 people, reduction by the second last cold period around 50 kya, and worse during the Last Glacial Maximum at 25 kya that separated humans basically into African, South(west) Asian, and East Asian groups that each share a certain number of traits (to avoid the outdated and unscientific designation ‘race’). Europeans derive from SW/S. Asian and Amerindians from E/NE Asians. However, we all go back to Africa.²⁹ Then, the question may be raised what original African speech was like.³⁰

LINGUISTIC RECONSTRUCTION

In the absence of written records, this question may seem impossible to answer. Our written records are only some 5000 years old, starting with Mesopotamia and Egypt. They are followed

28 Like my then 20-month-old son did in Nepal, taking a look around from a mountain pass: “so many blue mountains / Himalaya-yama”. Most of us will have our own experiences of this ability of small children.

29 In spite of some residual resistance from a few paleontologists and geneticists pleading multilocal origins in rear-guard skirmishes.

30 M. Ruhlen’s reconstructions (1994a) of early Homo sapiens’ language and his global etymologies, however, are still rejected by most linguists. They maintain that it is impossible to reconstruct a language that long ago. However, the assertion by traditional linguists that the comparative method is incapable of dealing with data before c. 6000–8000 years ago, which is an assumed and unproven time frame, is easily contradicted by the early dates for the generally accepted Afro-Asiatic (‘Hamito-Semitic’) family.

by other cultures centuries or even millennia later.

However, there are a few methods in historical comparative linguistics and, as I have recently proposed,³¹ also in historical comparative mythology, which allow us to approach this distant period. As Darwinians may readily understand, the comparison of available modern skeletons or of ancient retrieved materials quickly leads one to set up a palaeontological scheme that looks like a family tree of one's relatives and ancestors, in other words, a phylogenetic tree.

In biology, this shows how humans, apes and other mammals are interrelated and how they are related to other beings such as reptiles and fish, and beyond. Obviously, this involves a recurrent undertaking of detecting the last common ancestor of the various subgroups (species, etc.) involved, a particular ancestor that caused the shared innovation(s) found in all descendants. Further, it involves showing how various respective common ancestors are interrelated in still earlier times, even if the earliest forms of life may no longer be resolvable in tree-like fashion (Woese 1998).

The same procedure can be used for the establishment of the history of other human traits such as their genetic make-up, and for human cultural products such as languages, religions, rituals, myths, folktales, gestures, music, art, tools, and the like. Even manuscripts of certain texts (such as the *Mahābhārata* or *Rāmāyana*) can be studied, and their 'family' relationships (*stemma*) discovered – which one has been copied from another (with transmission mistakes as in genetic copying) – a field of study that has been perfected over the past 200 years in philology but is now being automated by computer, following biological precedents.

In this way, we proceed in the historical study of language. This proposal of comparing languages has been tried and tested for more than 200 years by now (William Jones 1786, Bopp 1816). We can compare, for example, some modern Indian languages and will quickly notice that some are closely related, while others are not.

In doing so, we can compare the array of sounds used, syllable and word structure, individual words, grammatical forms of

31 Witzel 2010 (in press); earlier: 2001: 45–62.

verbs or nouns, or the word order within a sentence (syntax). Some of these items are more useful for comparison than others. Mere similarity of words (and their meanings) across languages is commonly used initially to discover rough groupings (*see table below*); in other words, to suggest language families. However, due to many chance similarities (like Greek *theos*, Aztec *teo* ‘god’), this is not enough to firmly determine the existence of a language family and to show their intricate internal relationships. For this, we also have to establish *regular correspondences* between the sounds and the grammatical forms of the languages compared (described in some detail below). Both combined indicate common descent, a true phylogeny. On the other hand, comparisons merely of sound systems or word order does not help much in establishing language families as these features are easily borrowed, something that was understood early on in European languages. Indeed, regardless of origin, virtually all languages of S. Asia have the same word order in sentences (and share most of their sounds) now, but some (like Munda) did not have these features from their beginnings. Similarity of such features rather indicates mere past proximity, not patterns of inheritance from a common ancestor.

As an example of clearly visible patterns of unbridgeable differences that can be detected immediately by any one, note those of the numbers 1–10 in the three language families: Indo-European (Sanskrit, Hindi), Dravidian (Tamil), and Austroasiatic (Mundari).

	Sanskrit	Hindi	Tamil	Mundari
1	eka	ek	oru, ōr	miya’d
2	dvau, dvi	do	iraṅṅu, reṅṅu	bar-ia
3	trayaḥ, tri	tīn	mūṅṅu, mūṅṅu	api-a
4	caturāḥ/catur	cār	nāl(u), nāṅṅku, nālku	upun-ia
5	pañca	pāṅc	aintu, aṅcu	mōñē-a
6	ṣaṭ	cha	āṅṅu	turui-a
7	sapta	sāt	ēṅṅu	e(i)a
8	aṣṭau	āṭh	eṭṅṅu	iral-ia
9	nava	nau	oṅṅpatu, ompa/ōtu	arē-a
10	daśa	das	pattu, paḅṅtu	gel-ea

It is immediately clear that these numbers are unrelated in the four languages tabled here (except for the Skt–Hindi connection). The same non-relationship holds for their close relatives, e.g. respectively, Marathi or Bengali for Indo-Aryan, Telugu or Kannada for Drav., Santali or Sora for Munda, etc. Similarly, various fish, reptiles and mammals, or on a more recent level, the orang utans, chimpanzees and anatomically modern humans appear quite similar to each other, but are quite different once an informed biological comparison is carried out.

The rest, thus, is for the linguists to figure out: how the individual languages are interrelated and how they are related to their ancient ancestors, such as preserved in India in Vedic Sanskrit (of c. 1500 BCE) or the earliest Tamil inscriptions (of c. 150 BCE) recently published by Iravatham Mahadevan in collaboration with HOS.³²

Early on, one has done such a comparative study for European numbers and could determine the families involved even before William Jones' Calcutta speech in 1786 (cf. Lord Monboddo, 1773–1809) announcing the Indo-European family. Since then, and especially since Franz Bopp (1816), we have refined the Indo-European relationship at great length, so that if I were to meet an ancient Indo-European, he or she would wonder about my strange accent but would understand me; the same holds, of course, for the reconstructed Proto-Dravidian languages of south and central India.

It is useful to briefly evoke the method used in reconstructing the taxonomic tree of Indo-European, Dravidian etc. One of the first securely established items of Indo-European mythology, 'Father Heaven', is a good case in point (nominative and accusative):

Sanskrit	<i>dyāus pitā(r)</i> ,	<i>pitaram</i>
Greek	<i>zēus patēr</i>	<i>paterem</i>
Latin	<i>iu-ppiter</i>	<i>patrem</i>
Germanic	<i>tiu (+ Goth. fadar)</i> , <i>Engl. Tues-(day)</i> , (<i>+ father</i>)	
	thus: in reconstructed ³³ Indo-European:	
PIE	<i>*diēus ph2tēr</i> ,	<i>*ph2terem</i>
	Here certain sounds of Sanskrit, Greek, Latin, Germanic are	

32 Mahadevan 2003.

33 Reconstructed words are marked by an asterisk, thus **diēus*.

found in a regular relationship (p/f, e/a, etc.) that is normally met in (nearly) all other words of the languages involved. Such *lautgesetze* occur in a particular word of the same meaning in each of the related languages involved.

Superficial similarity of some words (like Greek *theos*, Aztec *teo*) does not constitute proof at all. (On the contrary, Armenian *hair*, Irish *athir*, English *father*, have fairly little in common when compared with the more ‘regular’ Latin *pater* [French *père*], Greek *patēr*, Sanskrit *pitā(r)* and Tocharian *pācar/pacer*, but they all are closely related by regular sound correspondences.)

Two similar-looking words such as English *heart* and Sanskrit *hṛd* ‘heart’ or German *kaufen* and Japanese *ka(f)u* ‘to buy’ are historically unrelated. It has often been said that one can find 50 words in any two languages of the world, which look somewhat similar and have a similar meaning. This is *not* proof of genetic relationship. Another archaic example:

	‘he/she/it is’	‘they are’
<i>Indo-Eur.</i>	*h ₁ és-ti	*h ₁ s-énti
<i>Sanskrit</i>	ás-ti	s-ánti
<i>Greek</i>	es-ti	-- (eisin)
<i>Latin</i>	es-t	s-unt
<i>French</i>	es-t	s-ont
<i>German</i>	is-t	s-ind
<i>English</i>	is	-- (are)

The proof of the pudding is not just this pervasive regularity (as we also see in biological reconstructions) but the fact that many of the more refined reconstructions of the later nineteenth century have subsequently been reconfirmed by missing links. This is similar to the famous archeopterix that is situated between reptiles and birds. For example, the *reconstructed* laryngeal sound *h₂* in *ph₂ter* ‘father’ that was not attested in any Indo-European language was later on, in the early twentieth century, discovered in written form(!) in documents in tablet form, when a previously not deciphered language, Hittite, was first understood. There, this sound, otherwise lost in all other Indo-European languages, is now seen as actually written. There are other cases, such as in the recently (1948) deciphered Mycaenean Greek that has preserved some pre-Homeric sounds. As in biology, each new discovery

brings about more filling in of the available reconstructions.

The result of such systematic comparison is the establishment of the ancestral Proto-Indo-European vocabulary and grammar. The vocabulary provides, via linguistic paleontology, some data about the life and time of the people speaking Proto-Indo-European: they already knew about the 'red' metal, copper, and had oxen-drawn wagons with full, heavy wheels. The combination of both facts points to the late fourth millennium for the parent language. There was some agriculture (barley, oats) but they heavily relied on pastoralism (horses, cattle, sheep, goats). Cattle (*pek'u*) was a term for riches (Latin *pecunia*). Grammatical analysis tells us that the horse is a relatively new acquisition (as indeed it is, c. 4000 BCE in Kazakhstan), and there was no horse-drawn spoke-wheeled chariot yet (invented only in c. 2000 BCE). The features of nature, plants and animals point to a homeland area with a temperate climate, close to a sea, with snow, birch ('the white one'), oak, wolf, bear, beaver, honey bees and their product, honey and mead. One therefore has thought of a location in the Eurasian steppe belt, perhaps in the Ukraine or Southern Russia but the exact location remains uncertain. Early loan words from neighbouring languages (Sumerian: copper, Caucasus languages: wine) point in the same direction. Tropical plants (palm tree, bamboo) and animals (tiger, lion) are absent.

If we then go on comparing the other (c. a dozen) major language families across the globe and try to establish their interrelations, we would ultimately reach the long lost language of the African Eve. Irrespective of the lost languages of her 'sisters' and of many of her direct descendants which have all disappeared without a trace due to bottlenecks in human descent and spread, some similarities in *all* surviving language families point in that direction (see below, for 'milk' and 'finger').

However, that work has not been done, and will not be done for a long time, even with the help of powerful computers, because of the complexities and huge amount of data of the c. 6000 languages to be evaluated. There is no computer program that can do so, yet, due to the ambiguities in meanings. English *queen* just means 'woman' in Gothic (*q^vino*), and English *hound* means 'dog' in German (*Hund*), while English *dog* means 'hound' in German (*Dogge*), etc. So, we merely get 'some kind of female', 'some kind of

dog’ – and these are the easy cases.

Long-range and deep-level comparison is actually opposed by ‘traditional’ comparative linguists who maintain that the noise-to-content ratio in going back further and further in time is too great to retrieve reliable information. Curiously, their time limit, some 6000 years, agrees with the limit set for Proto-Indo-European. We will see what can be established.

African and extra-African languages diverged after the exodus of early humans from East Africa at c. 75 000–65 000 years ago. This is a common feature in language development occurring when two communities are no longer in regular contact (such as the Dutch/Flemish and the South African (Afrikaans) version of Dutch, or even British and American English). Both the African and extra-African branches further developed in time, resulting in the currently assumed 12 major languages families, comprising some 6000 surviving languages. New Guinea alone has about a thousand, mostly little-studied languages, and India, depending on where one sets the bar between languages and dialects, can easily compete.

4. THE LANGUAGE OF THE AFRICAN EVE AND PAN-GAEAN MYTHOLOGY

Starting out from an unknown African language c.150 kya ago, there are some language groups in Africa that may be modern descendants: such as Bantu, a part of Greenberg’s Niger-Congo group, or Nilo-Saharan, and especially the archaic click languages of South Africa and a remnant in Tanzania (Hadza, Sandawe), whose genetic haplogroups, incidentally, are among the earliest in the pre-Exodus period.³⁴

John Bengtson, Vaclav Blažek and M. Ruhlen have actually proposed worldwide etymologies for a few³⁵ truly ancestral words. However, they could not and did not attempt the reconstruction of a complete language that would reflect this ancient most human linguistic inheritance. Examples include *maliq’a ‘to suckle, nurse, breast’ (Engl. milk), *tik ‘index finger, to point, one’ (Engl. in-

34 S. Tishkoff 2007

35 See summary in Ruhlen 1994

dex, in-dicate, Skt. diś, etc.), pal ‘2’ (Mundari bar-ia), or kaka ‘a relative’ (Hindi kākā),³⁶ etc.

The non-African rest of the world’s languages then must go back to the early emigrants of c. 75–65 kya which are attested archeologically as well (Tamil Nadu c. 75 kya, Australia 40–60 kya, etc.). Their language has not yet been reconstructed either, as this would depend on a clear description and individual reconstruction of all Eurasian and Australian (Sahul Land) languages, from which we still are a long way off.

However, there is an additional counter-argument, going much beyond the protests of the traditional comparativists. According to some scholars, it is not yet altogether clear when early *Homo sapiens sap.* could actually produce syntactically arranged proper speech.

Lieberman, for example, holds that this was possible only after c. 50 000 BCE.³⁷ But this applies only to fully vocalized speech. It is much too late if we accept that Australians moved into their continent between 40 000 and 60 000 years ago but already brought, as I have shown elsewhere, a particular ‘Southern’ (Gondwana) style mythology with them, as an offshoot of the out-of-Africa movement at c. 65 000 BCE (or according to some, even at 77 000 BCE). We have recently learned that even our ape relatives can

36 See Bengtson and Ruhlen 1994: Global etymologies <http://www.merittruhlen.com/files/Global.pdf>

37 Lieberman 2006. He maintains that early humans and Neanderthals could not produce basic vowels (such as a, i, u), but just the rather undifferentiated schwa vowel (ə) that is heard in the pronunciation of a in about, the e in bulletin or the i in ‘tangible.’ This clearly overlooks the fact that there are languages that use also other vowels than [a,i,u], such as r in Croatian *krk* and Sanskrit *vṛka*, or l as in English *bottle* (and that Kabardian in the Caucasus has been alleged to have no vowels, probably wrongly). One can produce words of the ‘Clingon’ type *tkx* (with the vowel *x*), Croatian *krk*, Czech *vlk*, which is enough for regular communication. If indeed early humans could only pronounce ə, that early language could transmit their thoughts and mythology just as well as current languages. Early human language thus may have looked quite different (for example, involving clicks) from the one Lieberman assumes for periods before 50 kya, though he thinks ‘that fully human syntactic and cognitive abilities were also present’ (Lieberman 2006: 59). This would indeed have been necessary for the development of both the Gondwana and Laurasian mythologies (Witzel 2010). For the emergence of anatomically modern human behaviour, see H. James in Petraglia and Allchin 2007: 204 sqq.

produce syntactically arranged sounds.³⁸

This argument of a late appearance of speech cannot simply be undone by pointing out some early cultural artifacts like tools, which could even be made without ‘proper speech’. However, the emergence of complex art using symbolic representation – the so-called artistic explosion of c. 40 kya – was not as sudden as usually maintained.³⁹ Symbolic use of art actually goes back to at least 90 kya (in Algeria, S. Africa).⁴⁰ Most importantly, the appearance of the closely related African and Australian mythologies speaks against such a scenario: how was it possible to transport the same original, complex (Pan-Gaeon) myths out of Africa around 75 kya and onward to Australia at c. 60–40 kya, without proper speech? The ‘family tree’ of genetic haplogroups, languages and mythologies would look as shown in the following, abbreviated table.

ORIGINAL AFRICAN		
Genetic lineages mtDNA L1–3 NRY A, B	Language families Nilo-Saharan Niger-Congo Khoi-San/Hadza, Sandawe	Mythology types Gondwana mythologies: African mythologies; Khoi-San mythologies
AFTER THE EXODUS, c. 75–65 kya		

38 See above, note 7:

<http://www.pnas.org/content/106/51/22026.abstract?etoc>.

39 An important study of cultural evolution, focusing on the symbolic explosion that occurred ca. 45K years bp:

<http://www.sciencemag.org/cgi/content/abstract/324/5932/1298>.

Overview account in Science Daily:

<http://www.sciencedaily.com/releases/2009/06/090604144324.htm>.

40 James Harrod, 2006; in detail: Researching the origins of art, religion and mind. <http://www.originsnet.org/glossmeth.html>.

mtDNA L3 □ M, N □ R NRY D, C	Andamanese, Tasmanian, Papua/ Melanesian; Australian languages	Gondwana: Andamanese, Tasmanian, Papua/ Melanes.; Australian mythologies
Expansion North- ward, c. 40 kya		
mtDNA NRY *F (F-S)	'Borean' = Dene- Caucasian (Basque, N. Caucas., Burushaski, Ket, Na-Dene (Apache etc.); Nostratic/Eurasiatic; Austro-Thai; Austroc: Austronesian, Austro- Asiatic; Tibeto-Burman, etc.	Laurasian mythologies of most of Eurasia, exc. for some Gondwana refuge areas (Toda, Semang, Aeta, etc.)
Immigration to the Americas		
NRY ABCD	Amerind languages; Eskimo-Aleut	Laurasian mythologies of the Americas

5. BOREAN AND NOSTRATIC

The original African and the 'Exodus language' split up in the course of time (after c. 60 kya) and this has led to about a dozen major language families, most of them outside Africa. The well-established language families included, by the mid-twentieth century:

- Nilo-Saharan (in the Sahel and Sahara belt of N. Africa)⁴¹
- Niger-Congo (including Bantu in Central, East and S. rica)⁴²

41 Bender 1996, Heine et al. 2000; Ehret 2001

42 Wiliamson and Blench 2000, Nurse and Philippson 2003

- Khoi-san in S. Africa and their distant relatives in Tanzania (Hadza, Sandawe)
- Afro-Asiatic, (or Afrasian, Hamito-Semitic) in the northern half of Africa and the Near East⁴³
- Indo-European in Europe, Armenia, Iran, N. India, Sri Lanka⁴⁴
- Uralic (including Finno-Ugrian) in Northern Europe and Siberia⁴⁵
- Dravidian (S. Asia)
- Altaic (Turkish, Mongolian, Manchu, Ewenki, including now also Korean and Japanese)⁴⁶
- Sino-Tibetan⁴⁷ (Tibeto-Burmese/Burman and Chinese)⁴⁸
- Austric⁴⁹ (including Austro-Asiatic and Austronesian)⁵⁰ in Central and E. India, Nicobar Islands, Burma, Malaya, Cambodia, Viet Nam; Malayo-Polynesian/Austronesian in Taiwan, the Philippines, Indonesia, Madagascar and the Pacific;⁵¹ in addition,

43 Ehret 1995

44 Beekes 1995

45 Décsy 1990

46 Scholarship tends to fluctuate between the two extremes every few decades; after the unifying positions personified by N. Poppe and S. Starostin vs ‘splitters’ in the sixties (G. Clauson and G. Doerfer), some have now started to doubt the very existence of the Altaic family again, see A. Vovin, in Osada and Vovin, 2003. Proof for Altaic is given by Robbeets 2005.

47 Now called Tibeto-Burman by some, notably van Driem (2005, 2006)

48 For a brief overview see van Driem 2006; his ‘Tibeto-Burman’ includes Chinese as a northern outlier. For a discussion of current views, see Sagart et al. 2005, Thurgood and LaPolla 2003/2007.

49 For Austric see van Driem 2007; Cf. also Benedict’s proposal of a Japanese-Austro-Tai family 1990. For Austro-Tai see Benedict 1975, for its relationship to Austroasiatic see Benedict 1976.

50 For a brief overview see van Driem 2006; Anderson 2001a,b (<http://www.fas.harvard.edu/~witzel/Andersonaamorph-rtf.pdf>), 2007, Diffloth 1976, 1994, 2001, 2005; on related genetics see van Driem 2006: 173 sq

51 Also including, according to some, Miao/Hmong and some other languages in S. China, such as Tai-Kadai, which includes Thai in Thailand. For their classification, see J. Bengtson 2006, Benedict 1976, and the multifaceted, divergent recent discussions in Sagart et al. 2005 (by Ostapirat, Reid, Sagart, Starosta).

Tai-Kadai⁵² or Austro-Tai

- Papuan (with some 700 largely still unexplored languages in New Guinea) and Melanesian⁵³
- Australian⁵⁴
- Amerind⁵⁵ (N. and S. America, excluding: the Na-Dene⁵⁶ languages of Alaska, Yukon and the Navajo-Apache of the southwestern US).

Only a few languages remained totally isolated, such as Basque,⁵⁷ the extinct Etruscan,⁵⁸ and Kusunda in the mountains of S. Asia; Ket⁵⁹ in Central Siberia, Ainu⁶⁰ in Japan and Sakhalin, the E. Siberian Chukchi-Koryak and Kamchadal,⁶¹ the Inuit (Eskimo)⁶² in N. America, etc.

The intervening stages between the assumed Exodus language and these families are yet to be established. Little work has been done on the Southern Eurasian/Australian groups and its later northern 'Borean'⁶³ extensions, as these themselves still are in the process of reconstruction by a handful of Long Range linguists.

52 Diller, Edmonson and Luo 2008. In India (Assam) we have: Ahom, Phake and Khanti.

53 See the recent update by Black 2006, Whitehouse 2006; Wurm 1972, Foley 1986, Pawley and Ross 1995.

54 Wurm 1977, Dixon 1980, 2002, Clendon 2006

55 Greenberg 1987

56 Enrico 2004

57 Hualde & de Urbina 2003

58 Bonfante 2002

59 Vaijda 2004

60 Vovin 1993

61 Fortescu 2005

62 Spalding 1992, 1998. Eskimo has been linked by Greenberg to the Eurasian (~Nostratic) family.

63 Sergei Starostin, see <http://en.wikipedia.org/wiki/Borean>; cf. H. Fleming, http://greenberg-conference.stanford.edu/Fleming_Abstract.htm.

6. INDO-EUROPEAN, DRAVIDIAN, AUSTROASIATIC, ETC.

In the Indian subcontinent, there are three major, or rather six, language families.⁶⁴ The three most important current language families of South Asia are Indo-European (here represented by Sanskrit and Hindi), Dravidian (Tamil), and the Munda group of Austro-Asiatic (here Mundari of Jharkhand), further Burushaski,⁶⁵ the relative of some Caucasian languages (not Georgian,⁶⁶ but Cherkes, Chechen, etc.⁶⁷), as well as Kusunda⁶⁸ and Nahali.⁶⁹ There also are Tibeto-Burmese and some small languages: Burushaski and Andamanese, and some residue languages such as Nahali, Kusunda and Vedda. A listing of their numbers for 1–10 immediately shows that these languages are not related (cf. list given above).

These three language families are ‘genetic’, just like Darwin’s phylogenetic ones, based on shared characteristics that lead back to a common ancestor for each of them which developed the typical Indo-European, Dravidian, Austro-Asiatic characteristics in question, as mentioned earlier.

The common ancestor of Indo-European can be reconstructed for c. 3000 BCE (with the use of bronze, oxen-drawn wagons). Similarly, Proto-Drav., with a corresponding time (use of metals),⁷⁰ and Austro-Asiatic, probably of a similar time frame, if we take into account the settlement of SE Asia by their linguistic relatives whose subdivisions include the Munda, Khasi, Mon, Nicobarese, Khmer, Asli, and Vietnamese.⁷¹ The subdivisions of the Munda family are subject to continuing discussion, just as similar refinements are

64 For an account of the language families of South Asia and their development as well as substrates, see Southworth 2005.

65 Berger 1998, Tiffou and Morin 1989.

66 Hewitt 1995

67 Nikolayev and Starostin 1994

68 Watters 2005

69 Kuiper 1962

70 See Krishnamurti 2003

71 <http://www.ling.hawaii.edu/austroasiatic/AA/pinnow-map.jpg>

still being made to the intra-family relations of Indo-European, especially the somewhat anomalous situation of Germanic (i.e. English), or the 'western' style (called Kentum) languages such as Tocharian, found far in the easternmost Indo-European area (Xinjiang), and Hittite and its close relatives in Anatolia.

Incidentally, the frequently heard opinion (in India, that is) that the Dravidian languages (Telugu, Tamil, etc.) belong to the same 'Aryan' language group as the North Indian languages (Hindi, Marathi, etc.) is based on the medieval fiction that regarded all languages of India as derivatives of Sanskrit, some closer some less close, with *deśī* (or *mleccha*) words interspersed. It is now also being pushed by those who regard the split between northern (IE) and southern (Drav.) languages as a ploy of British colonialism. This division is rather obvious to linguists and to objective observers everywhere (see the list of numbers, given above).

If it is then pointed out that the 'southern' languages share the same words as the 'northern' ones, this is certainly true but is due to cultural exchange, and not due to genetic inheritance from their parent languages (Proto-Indo-European and Proto-Dravidian). In other words, if we want to use the biological terms: it is due to something like epigenetic influence, not common (phylogenetic) descent. The epigenetic property shared in these cases includes certain words, spread of certain 'South Asian' sounds (such as retroflexes: *ṣ*, *ḍ*, *ṇ*), word order, etc. (discussed later on).

As cases in point, words such as *dharma*, *karma*, or recent ones like *viśvavidyālay*, *ākāśvāṇī*, *dūrdarśan* are derived from the common Pan-Indian 'cultural' language, Sanskrit, and they are not native in the languages concerned. These words, in fact, are not even 'native' in modern Indo-Aryan languages, where Vedic *dharma* became *dhrama* in the northwestern Dardic languages; Old Beng. has *dhāma* 'religious conduct'; Hindi has *kām-dhām* 'work, business'; Old Singhalese has *dama* 'religion', cf. also Sindhi-Kacchi *a-dham* 'tyranny'. In short, the form of modern *dharma* is a direct medieval loan from Sanskrit and not derived from the natural development of the Indo-Aryan languages concerned. In the same way, it has been loaned into Dravidian languages.

Similarly all over Europe, we find medieval and modern loan words derived from Latin or Greek, like *university*, *religion*, *radio*, *television*. They are entirely non-native in modern Indo-European

languages just as they are in the Uralic languages Finnish, Estonian or Hungarian, languages that are entirely unintelligible to the majority of neighbouring Europeans, just as Dravidian is to Indo-Aryan speakers.

In sum, the three major language families of S. Asia have their own separate origins and their own individual development, continuing to this very day. The situation is again comparable to Europe, where we mostly have Indo-European languages but also a strong area of Uralic and an old remnant language, Basque, as well a few more recent newcomers: Maltese (Arabic), Turkish (Altaic) and Kalmyk on the lower Volga (Mongolian, e.g. Altaic). Nevertheless, most or all of the peoples speaking these languages regard themselves as Europeans and share most of the common European culture. Similarly, this is the case for India.

However, there are a number of residue languages that do not belong to these three major families.

7. RESIDUES: A COMPLEX PREHISTORY

To begin with, it has to be understood that every language that we speak retains earlier stages of its individual development, in quasi-archaeological layers, much as humans retain some feature of stages in the past, such as the appendix or tailbone.

To give a well-researched parallel: in English, these layers include, first, the recently acquired Indian terms (guru, pandit, yoga, bungalow, verandah, curry, punch, pajama, etc.). Below that level there are Dutch sailors' words (boss, mate, dike, etc.) then the ever-increasing multitude of 'learned', originally humanistic Graeco-Roman words of the Renaissance. Before that, the equally numerous Norman French words (after 1066 CE, like river, mountain, forest, ville, court, castle, army, royal, grand, beauty, etc.), and some Viking Old Norse words (egg, -vik, -by, etc.). Only then we get to the layer of ancestral Anglo-Saxon words that form the core of English (words like father, mother, water, horse, cow, milk, to go, to do, numbers and particles). They have come over from northern Germany and western Denmark with the Saxons, Angles and Jutes. Even before that, the western Germanic language Old Saxon had experienced a strong Roman input (street, mile, cellar, castle, etc.). But there is also a non-Indo-European residue of

some 30% of the Saxon (and Germanic) vocabulary that belongs to a pre-Indo-European, unknown 'North Sea' substrate, with words like sheep, eel, roe, boar, sour, lentil, land, delve, prick.⁷² We can detect similar subsequent historical levels in Indian languages.

In addition, it must not be forgotten that Old English (Anglo-Saxon) is clearly an 'invaders' language' and we therefore find remnants of substrates preceding them in Britain: local British Romance/Latin (-chester/-cester, channel, etc.), Insular Celtic (dune, London < Lugdunum), and even of an unknown northern British substrate (Pictish). The same situation obtains within the earliest Sanskrit, that of the *gveda*, which has some 380 substrate words⁷³ from unknown language(s).⁷⁴

Further, we can well reconstruct the subsequent developments, say from Indo-European and Vedic to Classical Sanskrit, and then onward to Hindi/Bengali, etc., for example (*asterisk indicates a reconstructed form; > indicates 'developing to', < 'derived from', ~ 'related to'):

Vedic *hasta* 'hand' > Pali *hattha* > Hindi *hath*

Vedic *akṣi* 'eye' > Pali *akkhi, acchi* > Hindi *āṅkh*

Ved. Skt. *hṛdaya* > Pali *hadaya*, Prakrit/Old Hindi *hīa* (now:

Persian *dil* < Avestan *zard* ~ Ved. *hṛd-*)

IE **k'erd* > *śrad* in *śrad-dhā* 'to put your heart into something'
= 'to trust', derived from another word for 'heart.'

Still, we always come across some words that stand out from this general IE/IA scheme as they do not reflect the expected, regular forms. For example, in Hindi, Tamil, or English for that matter, you do not expect a modern word like Mfume (an African American politician) Mbow (the name of the former Director General of Unesco), or Nkrumah (the name of the first president of Ghana) as mf-, mb- or nk- are not allowed as the beginning of words. The same applies to dialect words imported from neighbouring areas that do not fit the local language (Vedic Sanskrit).

The same pattern as observed above in English loan words from Dutch, Celtic, French, Latin, etc. can be seen in the influences of loan words in South Asia. This process has been going on at least

72 For these 'North Sea area' words cf. Witzel 2003.

73 Kuiper 1991

74 Witzel 1999

since the Persian conquests of the Greater Panjab in 530 BCE: band (not bandhī), karṣa (not *kar a), pustaka (not *pu aka!) now are common Skt. words. Later on, Greek words were taken over after Alexander's invasion: mel , kalama, hor , jam tar; and still later, Turkic/Mongolian ones: bahadur, begum, horse terms ending in – ha such as kho g ha (< Turkic qo gur 'red'), innumerable Persian, Arabic, English and some Portuguese (kamiz, sabun) words, and now sushi and kungfu... There also is the multitude of early loans from Dravidian and Munda.

A list of loan words in Indian languages is given below, from modern times down to the oldest forms of Sanskrit (and also Dravidian). It roughly follows the many historical levels of 'foreign' and substrate influences; for illustrative purposes, a list of the (sometimes corresponding) levels of loan words in English is juxtaposed.

TABLE COMES HERE

However, though the study of such loans is very important for cultural history, all of this deals, in the end, only with incidental influences on Indo-Aryan. Instead, taking seriously the pre-Indo-Aryan remnant (substrate) words, found from the *gveda* onward, like the c. 30% in English, we can detect a growing number of words that go back to none of the major language families of S. Asia, such as Indo-European, Drav., Munda etc., which we are in the process of collecting, at our SARVA website.⁷⁵

Indeed, South Asia is a very complex area, with 5 or more language families and a lot of inferred remnants of lost languages and families, which is not surprising. For comparison, in Europe, we have clear data on languages that precede the Indo-European ones: not just Basque, but a whole array of fragmentary languages from Spain to Turkey, and from Scotland to the Alps.⁷⁶

The obvious corresponding ones in India are: Burushaski in northernmost Kashmir, just below the Pamir passes in Hunza. Usually it is treated as an isolate, a remnant of a lost family that has no relatives left anywhere else. However, some more ‘adventurous’ linguists, such as John Bengtson, have shown that it is related with Northern Caucasus languages and Basque (on the French/Spanish border). This Macro-Caucasian language family reflects the remnants of the early move out of western South Asia/Iran into West Asia and Europe at c. 40 kya.

A similar case is that of Andamanese, spoken on some islands such as Great Andaman, Little Andaman and Sentinel. A few groups remain: Andamanese, Jarawa, Önge, (the still unstudied) Sentinel, and the neglected Shompen on the Nicobar Islands. They were regarded as isolated until the late J. Greenberg united them in a new language family (Indo-Pacific) along with Papuan and Tasmanian. Recently, and rather surprisingly, another language in the hills of Nepal has been added.⁷⁷ This is Kusunda, the language of one of the hunter-gatherer groups in Central and Western Nepal.

75 <http://www.aa.tufs.ac.jp/sarva/>. This is a long-term project undertaken by F. Southworth, M. Witzel and D. Stampe. It is hosted by the Institute for the Study of Asian and African Languages and Cultures of the Tokyo University of Foreign Studies.

76 Sverdrup 2002: 117, and other contributions in this volume

77 Whitehouse et al. 2004

Long pronounced dead, it was recently rediscovered by my friend B.K. Rana,⁷⁸ and has been described by him, and more recently in great detail by the late D. Watters (based on three speakers!)⁷⁹

AN OVERVIEW OF SUBSTRATES

We have by now inferred substrates that are not directly attested, unlike the still spoken languages Andamanese, Kusunda (Nepal), or Vedda⁸⁰ (Sri Lanka). Substrates include those found New Indo-Aryan languages: Khovar (Chitral in NW Pakistan);⁸¹ a strong substrate in Kashmiri (some 25%)⁸² along with strange vowels and an aberrant syntax; Tharu in the Siwalik Himalayan foothills of India and Nepal,⁸³ inscriptional evidence in the Kathmandu Valley,⁸⁴ materials relating to the Indus period of the Greater Panjab⁸⁵ and Sindh (Meluhhan),⁸⁶ many agricultural words of the Ganges Plains ('Language X'),⁸⁷ words in Bhili⁸⁸ and some 25% of the Nahali vocabulary (Central India),⁸⁹ Vedda of Sri Lanka, and in some Nilgiri hill languages.⁹⁰

There will be many more, but such substrates have not been discovered as little work has been done to investigate them. Who knows what may still be discovered in the nooks and corners, in the mountains and jungles of the subcontinent? There is much work to be done before many of these 'tribal, jungli' languages disappear.

78 Rana 2002

79 Watters 2005; his dictionary is found at http://www.aa.tufs.ac.jp/sarva/materials_frame.html.

80 de Silva 1962, Witzel 1999

81 Kuiper 1962

82 L. Schmid 1981

83 Witzel, unpublished pilot study

84 Witzel 1999

85 Witzel 1999, Kuiper 1991

86 Witzel 1999

87 Masica 1979

88 Koppers 1948

89 Kuiper 1962

90 Zvelebil 1990

The Indian Institute of Language Studies in Mysore (<http://www.iils.org/>) and the Indian Anthropological Survey (<http://www.ansi.gov.in/default.htm>) do some of this work, but there is need for much more. These languages are precious testimonies to early human history, and each language lost is a loss for all of humanity. Political and administrative hurdles should not be a bar.⁹¹

However, a so far almost completely overlooked field is that of the external relationships of Dravidian that covers all of South and much of Central India. Many historians and some linguists assume that Dravidian has 'always been present in India', especially so in the southern subcontinent. Some assume a relationship with Uralic. However, Vaclav Blažek (Brno, Czech Republic) has recently pointed out a promising number of words in Dravidian that seem to stem from a lost Australian substrate⁹² in India. Applauded by a specialist in Australian Languages (Paul Black, at our 2006 Round Table),⁹³ Blažek has derived from Australian the numbers 1, 4, and words for eye, tooth, neck, shin, knee, feces and urine, etc. All of them belong to the basic vocabulary of any population (Swadesh list).

If his comparison holds up, we would have a very deep substrate in Southern India (and beyond?) that existed before the movement of Proto-Australian into the Australian continent. This must have been the case well before the Ice Age (c. 20 kya), possibly as early as c. 65 kya: note their otherwise rare retroflex sounds, below, and new, concurring archaeological and genetic finds in Tamil Nadu (Petraglia et al. 2007; Wells 2002; Majumder & Balasubramanian 2009; Friese 2009). Proto-Australian speakers would then have been present in India during the initial Out-of-Africa settlement, moved along the shores of the then exposed Indonesian Sunda Land subcontinent and crossed over by boat from Timor, and entered Australia by c. 50 000 BCE when the first human remains are found at Mungo Lake.

91 See now the various promising international projects to record such languages (and texts): Hans Rausing Foundation, U.K. (<http://www.hrelp.org/languages>); on oral traditions see (<http://www.oralliterature.org/>); cf. Witzel 2010

92 See Blažek 2006

93 Cf. Black 2006

This is important: it would indicate, again, that South Asia has seen many levels of languages and their populations ever since the immigration from Africa, some 75 kya. The Australian substrate in Dravidian is obviously expected, as it has been found in genetic evidence as well.⁹⁴ However, the Australian connection and other substrates in Dravidian have hardly ever been researched seriously.⁹⁵

It must also be pointed out that, due to pathway inheritance from the Middle Ages and as early as by Patañjali (150 BCE), there has been the prominent idea of 'local' *deśī* and 'barbarian' *mleccha* words in Sanskrit; due to this pathway dependency, virtually no work has been done in India on substrate languages, and even historical comparative linguistics as such is little developed in India.⁹⁶

However, in order to understand more of early South Asian history, we need such studies. There is a pressing need to get young Indian graduate students into this field and, first of all, to produce etymological dictionaries of the major languages of S. Asia. So far, we only have the exhaustive and up-to-date scientific etymological dictionary by M. Mayrhofer (Vedic and Classical Sanskrit), the merely descriptive New Indo-Aryan one by R. Turner (1966, CDIAL) and the equally descriptive Dravidian one by Th. Burrow and M. Emeneau (1960/1984, DEDR). Both the latter are not of a high standard as they do not discuss individual word formations (Turner) and do not even provide reconstructions (DEDR). Instead, they provide just lists of words somehow related to each other. Further, generally the Tamil form is taken as being close to

94 Wells 2002, in collaboration with Dr Ramasamy Pitchappan, who studied the NRY of three populations in the Madurai region, where he found a person with 'pre-Australian' genetic traits (in: Wells et al. 2001; Friese 2009). Note also the PBS film: The journey of man: a genetic odyssey; see: <http://www.shoppbs.org/product/index.jsp?productId=1402989>. See now Majumder and Balasubramanian 2009.

95 Nothing in Krishnamurti; see, however, for a handful of words in the Nilgiri languages (Zvelebil 1990) that seem to be of unknown origin.

96 Such as in the off-mainstream writings of Misra 1992 and in Bryant & Patton 2005, all of this after a promising beginning by scholars such as Suniti Kumar Chatterji (1926, 1974) and others.

Proto-Dravidian, due to its frequently archaic form.

There are only rudimentary etymological dictionaries for Bengali⁹⁷ and Marathi⁹⁸ that do not go back beyond Sanskrit and merely list *deśī* words; there also is major recent one for Tamil.⁹⁹ Obviously, linguists studying the substrate words could do useful work alongside geneticists and archaeologists (e.g. for Nihali, Kusunda), but this is just not happening.

While all these additional and substrate language families are genetic ones in their own way, just like Darwins's phylogenetic ones, they cannot be brought into any simple Indo-European-Dravidian or expanded Indo-European-Drav.-Munda-Tibeto-Burmese scheme. Instead, they reflect other, wider families, such as Macro-Caucasian (Burushaski), Indo-Pacific (Andamanese, Kusunda), and beyond, which are lost families.

I still wonder about the linguistic connection between the isolated languages Nahali (in Central India) and Ainu¹⁰⁰ (in northern Japan and Sakhalin). Both share very similar words for 'fire, dog, monkey', among others. This connection may eventually also be reflected by the ancient genetic NRY chromosome haplogroup D that has so far been spottily found in N. and S. India, but not yet among the Nahals. This would again reflect the early settlement of much of S. and E. Asia. The Nahali language has some of the earliest substrates in India, some 25%, below Indo-Aryan, Drav. and Munda levels. Haplogroup D has been driven into the nooks and corners of the continent by later spreads. It still occurs in Tibet, the Andamans, Japan and in some Indian tribes, the Rajbamshi on the India Bengal/Nepal border, the Kurumba in the Nilgiris.¹⁰¹

The substrate languages mentioned earlier have exercised a certain amount of influence on all later-arrived or later-derived Indian languages. Notable is the curious case of the ancient retroflex

97 Sen 1971

98 Tulpule 2000

99 The multi-volume one *Ñ namuttan T van yan et al.* 1985-, however, includes many fanciful derivations.

100 Witzel 1999; Kuiper 1962; Bengtson 2006

101 Thangaraj 2003

consonants (as in *kūṭa*, *iṣṭa*, *heḍa*, *piṇḍa*, *daṇḍa*, *gaṇa*, see below). Unexpectedly, they are most prominent in the Hindukush/Pamir area but have affected languages in the rest of the subcontinent, even Munda. Curiously, retroflexes also are typical of aboriginal languages of Australia.

A few words are in order about the historical development from the early forms of the ancestral IE/IA, Dravidian and Austro-Asiatic ‘mother tongues’ down to their modern forms, such as Hindi, Tamil or Mundari.

8. DEVELOPMENTS FROM VEDIC SKT. TO MIDDLE AND NEW INDO-ĀRYAN

Old Indo-Aryan includes Vedic, Classical and Epic Skt.; Vedic developed to the more popular Middle Indo-Aryan languages (*Prākṛt*, *Pāli*, *Ardha-Magadhī*, *Māharāṣṭrī*, etc.). These, starting in mid-first millennium CE, gradually developed into the current New Indo-Aryan (NIA) languages. Starting in the Northwest we have Dardic (Kashmiri, Kalasha, Khowar, Kohistani, etc.), Panjabi, Dogri and Sindhi, the multitude of Pahari languages (including Garhwali, Nepali, etc.), the eastern branch with Maithili, Bengali, Assamese and Oriya; the central languages (Bhojpuri, Awadhi, Hindi/Urdu, Braj, various Rajasthani languages, etc.), and finally, Gujarati, Maharashtra with Konkani, and Sinhala in Sri Lanka.

There have been regular developments from the ancient form of Indo-Aryan (Vedic Skt.) down to the modern languages. They follow certain patterns, such as simplification of double consonants, as seen in *hasta* ‘hand’ > Pali *hattha* > Hindi *hath*, or in *akṣi* ‘eye’ > Pali *akkhi*, *acchi* > Hindi *āñkh*. Usually, single initial consonants are retained while those in the middle of words are ‘weakened’ (from -t- > d > ḍ > zero, etc.) and finally disappeared, as in *hr̥daya* ‘heart’ > Pali *hadaya* > Maharashtra, Old Hindi *hīaa*.

However, some Indo-Aryan areas have very different, ‘strange’ developments, such as *guru* > Kashmiri *gōr*, genitive *gur*’s, or Vedic *aṣṭau* ‘8’ that is retained in the conservative Kalasha *aṣṭ* (Chitral, NW Pakistan), while in the same Dardic group, it became Kashmiri *ōṭh*, which is closer to Hindi *āṭh*.

Other major developments include the complete restructuring of the Skt. nominal and verbal system, from a flectional system (*guruḥ*, gen. *guroḥ*) to a quasi-agglutinative noun declension (*guru-k*, *ke*, *k*), and Skt. *asti* / MIA **asati* 'he is' > Hindi *hai*, or *cal-th* < *cal-* + *sthitaka*, etc.

Except for the frequently archaic northwestern (Dardic) languages, most New Indo-Aryan languages follow these patterns, so that modern words and noun or verb forms no longer resemble the Skt. ones, the frequent loans from Skt. (like *dharma*, *karma*, *dardar an*) obviously excepted.

9. DEVELOPMENT OF DRAVIDIAN > TAMIL, ETC.

Similarly, regular developments can also be observed in the other Indian language families.

The Dravidian family is usually divided into three branches: a northern one with Brahui (in Baluchistan), Kurukh (Oraon) and Malto; the southern central one with Gondi, Telugu; a central one with a host of small languages (such as Kolami or Parji); and a southern one with Tamil, Malayalam, Kannada and Tulu.

Most developments from PDrav. to the individual languages are regular, such as Drav. *p* > Kannada *h* which occurred, as the inscriptions show, only around 900 CE. The Old Tamil of the Sangam texts and inscriptions (from c. 2nd century BCE onward)¹⁰² too has changed considerably over time, in spite of the relative conservative nature of Tamil consonants, vowels and word forms. For example, the sound *h*, a glottal fricative, is still preserved by the Old Tamil *ytam*, but soon disappeared in all of Drav. Or, the alveolar consonant *ç* changed in many languages into a continuant, the trilled *ç*.

Further, the reconstructed Proto-Dravidian and Old Tamil have only word-initial tenues (voiceless, surd) consonants (*k*, *t*, *p*) that can become 'weakened' in pronunciation to *h/x*, etc. in internal positions.¹⁰³ However, some languages such as Kannada and Telugu have them at the beginning of words,¹⁰⁴ and this is

102 See Cardona and Jain 2003

103 Mahadevan 2003

104 Krishnamurti 2003: 144 sqq.

even reflected in old Sanskrit loan words (such as in Skt. (b hu-) da a 'arm', Malayalam ta a, or Skt. ga a 'joint', Tamil ke ai 'ankle,' Kannada ga u 'knot of cord', Tulu ga , ga u, Telugu ga u, ga a. This raises the question as to exactly which branch of Drav. the early Vedic loans derive and when the change from k > g took place in some Drav. languages. Among other typical features of Drav. are the three types, clearly distinguished, of dental, alveolar and retroflex consonants (such as n, n and).

In this language family, too, there are 'strange' local developments, especially in some of the smaller languages such as Toda that is spoken high up in the Nilgiri Hills. Here, the development of *p has gone one step further, it has disappeared at an unknown point in time: Toda ĩ 'the dead,' ĩ 'the world of the dead' (n 'place'), ĩ o.w 'the god of the dead' belongs to the following Dravidian words: Tamil pi am, pi a 'corpse, carcass, disembodied soul, devil, spirit'; Malayalam pi am 'corpse, dead body of animals'; Kannada pe a, he a 'corpse, carcass'; Tulu pu a 'corpse, dead body; dead, inactive; pe a 'corpse.'¹⁰⁵

The aggregate effect of many such local Toda developments is best visible in complete sentences, such as: a koyš ũz, iĬ , pīs, it ĩyi 'For that reason one should do the bow ceremony in the fourth month (of pregnancy)' – which remains unintelligible and undecipherable for all other Dravidian speakers. But, as is obvious from the example of in 'dead', these words too are regular developments from older forms of South Dravidian.

The oldest states of Dravidian, the reconstructed Proto-Drav., can supply interesting insights into the thought and culture of that period. For example, 'pestle' ul-akk-, uram-kal may be a compound of DEDR 665 ur 'to grind' and 1298 kal 'stone',¹⁰⁶ as seen in DEDR 672: Tamil ulakkai 'pestle', Malayalam ulakka 'pestle for pounding rice', Toda was k 'grainpounder', Kannada olake, onake, onike 'wooden pestle for pounding rice and other things', etc. The word may have been loaned early into Vedic Skt. as ul khala- 'mortar' (in late¹⁰⁷ gveda, 1.28, where it is made,

105 Statistical discussion by Krishnamurti 2003: 132 sqq.

106 And Kor. hina id., Telugu p n gu, p ngu 'corpse, dead body, carcass,' Ga. p nige 'corps', Kon a p ngu, id. Kuwi p li , p nugu, pinugu id.

107 Krishnamurti 2003: 8

however, of wood).¹⁰⁸

On the other hand, in establishing Proto-Drav. etymologies, one should steer clear of items such as king, state, palace, fortresses surrounded by moats, etc. at c. 3000 BCE in a Neolithic, pre-agricultural and pastoral society.¹⁰⁹ To do so would be to commit the same, well-known mistake as when reconstructing ‘emperor’ from the modern Romance languages (empereur, imperatore, etc.) for their ancestral language Latin, when Classical Latin *imperator* still had the earlier, republican meaning ‘army leader, commander’.¹¹⁰

Among the typical characteristics of Drav., different from Indo-European/Indo-Aryan, are the following. Unlike the inflection of Indo-European (change of word stem, ‘fused’ endings and stem of words), Drav. is an agglutinative, that is the endings are merely added to the word in question; prefixes (as in IE/IA) and infixes (as in Munda) are not allowed. Unlike the three genders of Indo-European (male, female, neuter), there originally were probably four: singular: male human, non-male human, animate non-human, non-animate (which resembles Burushaski, below); the details of reconstruction, however, are still under discussion (Krishnamurti 2003: 210 sqq). For example in Proto-South Dravidian I, we have: *awan ‘he’, *awa, ‘she’ *atu ‘it’, *awar ‘they (m.)’, *away ‘they (f., non-human)’. The pronoun ‘we’ has inclusive (‘we all’) and exclusive (‘we, our group’) forms, for example: Tamil/Malayalam *n*, Telugu *manamu*, Kannada *n m*, ‘we all’ :: Tamil *y m*, Telugu *mu*, Kannada *m* ‘we, our group’.

Verbs originally distinguished between past and non-past, and all such verb forms also have negative forms. A single verb was positioned at the end of a sentence; however, in more complex sentences, it was preceded by non-changeable gerunds ‘having done this, having done that, ...’ (like Skt. *-tv*, *-ya*).

108 RV 1.1-50 and book 10 are late additions to the corpus.

109 See Turner, CDIAL 2360, (further H.-P. Schmidt, EJVS 16, 2009).

110 As reconstructed by Krishnamurti 2003: 7-8

10. DEVELOPMENT OF MUNDA, TIBETO-BURMESE, BURUSHASKI, ETC.

Munda

Munda languages represent a difficult case, as we have word lists and texts only for the past two centuries or so, and only a few words from Sanskrit texts of earlier periods.¹¹¹ Their close relationship with SE Asian languages¹¹² leaves an Indian or a southeastern Asian origin open. However, Patricia Donegan and David Stampe have recently argued that the SE languages diverged from the Indian Austroasiatic type (Munda) and changed structures, due to sentence accent which favoured either preservation of the initial or final sounds of a word and, while retaining the old prefixes, also originated the development of suffixes.

Similarly, Northern Munda (Santali, Mundari, etc.) diverged from the more archaic but little studied S. Munda (Sora, Gta, etc.). N. Munda was heavily influenced by the surrounding Indo-Aryan languages. It changed word order and individual structures.¹¹³

As we have written sources for this language family only for the past 200 years, the ancient forms of words and grammar and their historical developments are only accessible by reconstruction. Unlike in IA/IE or even in Drav., this is a work in progress.¹¹⁴ In addition, some of the more conservative Southern Munda languages such as Gta have hardly been described¹¹⁵ and closer comparison with the relatives of Munda in S.E. Asia (Mon, Khmer, Vietnamese) is required, a little found scholarly expertise. However, our reconstructions are sometimes borne out by attestations of single words in the older texts of South Asia. Thus, the Saora tribe is known in Skt. and even in late Vedic Sanskrit¹¹⁶ as *Śabara*.

111 In his draft, Krishnamurti even had ‘emperor’, until we (F. Southworth and I) alerted him. The other ahistorical assertions (above), however, remain.

112 For overviews, see <http://www.ling.hawaii.edu/austroasiatic/>

113 First in a detailed comparative study of Kharia, by Pinnow 1959.

114 See P. Donegan and D. Stampe: <http://www.ling.hawaii.edu/austroasiatic/AA/rhythm2004.pdf>.

115 Pinnow 1959; Anderson 2001, 2007, many details on the website of David Stampe: <http://www.ling.hawaii.edu/austroasiatic/>

116 Overview now in Gregory Anderson, *The Munda languages* 2008

The Munda languages are divided into a Northern and a Southern branch. The first is found in Jharkhand and surroundings (Bengal, Orissa): Santali, Mundari, Ho, Asuri, etc. with an outlier, Korku much farther west along the upper Tapti River; the Southern branch (Kharia, Saora, Gutob, Gta etc.) covers the border of Orissa and N.E. Andhra.

The linguistic character of this family differs entirely from the IA/Drav. ones. Originally, words were not formed by suffixes but mainly by prefixes and by infixes; that is, a consonant (mostly –n–) was inserted to change the form and meaning of a root. Though IE/IA also makes prolific use of prefixes (as in English be-have, be-get, be-moan, fore-go, for-get), this is most typical of Munda (but not found in Dravidian).

A number of prefixes are still actively used in Munda and Khasi, such as ab-/ob-, d -, etc., for example, Kharia su'd 'to be wet', o-su'd 'to wetten', Sora jum 'to eat', ab-jum, ajjum 'to feed', Khasi asam 'to drill', k-sam 'to insert nails'.

Most prominent are the actively used infixes, such as n-, –b-, –p-, which form nouns or causative/reciprocal verbs: Santali dapal 'to cover', da-na-pal 'cover', Sora gad 'to cut' g'-na-d- n 'piece', Santali dal 'to beat', da-pa-l 'to fight each other', Sora bato 'to fear', ba-b-to 'to frighten'.

We thus get the following words for 'bird': Kharia khon-the 'bird', Sora on-tid- n, tid- n, which has been loaned into Skt. as a-kuntaka, a-kuntal 'little bird'; a-kuna 'bird' (gveda), Kunti, Kunt, a-kuni (tribe and person).

Munda sentence structure is even more complex,¹¹⁷ in part caused by falling sentence accent with the original word order (subject-verb-object [SVO]) preserved in Sora d-m l-tiy-da-i -da-e 'he (-e) does not (d) want (m l) give (tiy) rice (dar) to me (i).' As in Drav., there is an inclusive and exclusive pronoun 'we.'

As mentioned earlier, the character of Munda has shifted, due to sentence accent or perhaps also under the influence of Indo-Aryan and Drav.(?) remarkably, for example, from an SVO word order in a sentence (as in Kashmiri) to a SOV word order. This would hardly be visible if one would only compare Munda materials: one has to take into account also the other Austroasiatic

languages: Khasi (Shillong hills) and the S.E. Asians ones, from Nicobarese to Vietnamese.¹¹⁸

TIBETO-BURMESE

This is a very large family of some 250–300 languages spoken in S.E, Central and East Asia.¹¹⁹ The ones of interest here are all along the Himalayan belt: in northern Kashmir the very conservative Ladakhi, various other southern Tibetan dialects (including Sherpa, and Dzonkha of Bhutan); Central Himalayan languages such as Magar, Gurung, Tamang, Newari, Rai, Limbu and Lepcha; and further east Bodo, Naga, Meitei (Manipuri), etc.

The earliest source of the Tibeto-Burmese language family is generally held to be Old Tibetan (along with Zhang Zhung). However, one has to take into account many names in the Kathmandu Valley, some nouns and names recorded in Sanskrit inscriptions, c. 450–750 CE. They are clearly of Tib.-Burm. type, though they do not represent a predecessor to current or medieval Newari.¹²⁰ The latter is first recorded in a land sale of c. 1000 CE.¹²¹ Tibetan is recorded a few centuries earlier, from the seventh c. onwards.¹²² Also, there are texts in pre-Hindu time Manipuri (in Meitei, c. 17–18th c. CE),¹²³ as well as Naga, Bodo, etc.

A reconstruction of Tib.-Burm., though attempted in the 1970s¹²⁴ is still in progress – due to the ‘compressing’ factors at work, resulting in monosyllabic words. For example, Old Tibetan *brgyad* ‘8’ (fairly well preserved in modern Ladakhi *rgyat*) has now become (g)y in Lhasa Tibetan.

Several models for Proto-Tib.-Burm. and also for the original

118 Pronouns at the head of a sentence; see Donegan and Stampe 2004

119 See Sino-Tibetan Etymological Dictionary and Thesaurus (STEDT) project, <http://stedt.berkeley.edu/html/STfamily.html#TBlg>.

120 Witzel 1980

121 Kölver and kya 1985

122 Note especially the non-canonic Dun Huang documents.

123 Unfortunately, many original manuscripts were burned recently, along with the Imphal library, due to political strife.

124 Benedict 1972; Matisoff 2003

homeland have been proposed.¹²⁵ The original form of this language family seems to have had a much more complex verbal system (with multiple suffixes) that is in part preserved in some Eastern Nepalese languages;¹²⁶ the few verbal affixes used in Old Tibetan point to another set of formations.¹²⁷

Tib.-Burm., however, has been spoken in areas bordering the northern parts of S. Asia since the Atharvaveda (c. 1000 BCE). One stanza mentions, in passing, the Kir ta girls collecting herbs, much as they do today in the Nepalese hills, and living in caves.¹²⁸ The term Kir t nowadays refers to the eastern Nepalese tribes, the Rai and Limbu. However, the Nepalese Goplar jav vali chronicle¹²⁹ of c. 1380 still refers to them as former inhabitants and kings of the Kathmandu Valley; before the arrival of the Licchavi dynasty in early CE. The Atharvaveda and the Maitr ya i Sa hit of the Yajurveda necessarily place them still further west,¹³⁰ in the H.P. area. Though we cannot be sure about the language these Kir ta spoke then, this western location reinforces the idea that the Himalayan belt may have been the homeland of this language family.

BURUSHASKI

Burushaski, spoken in Hunza in northernmost Pakistan, made its entry into linguistic and general consciousness very late. The

125 See summary by van Driem 2005, 2006 (including also genetics).

126 Originally thought to have been borrowed from the Munda languages, Benedict showed that these inflectional forms were part of the Proto-Tib.-Burmese structure.

127 Hill 2003

128 kairatik kumarik Paippal da Sa hit 16.16.4, aunaka Sa hi 10.4.14, kail ta 8.2.5; VS 30.16 has them as living in caves; cf. also the popular form Kil ta (Pañcavi a, Jaimin ya, atapatha Br hma a)

129 See Bajracharya and Malla 1985

130 A location still further west, in the Kashmir area, is seen in Hsüan Tsang's Hsiyuki (c. 600 CE), who knows of them as Kilito, as a people in Kashmir who had their own king shortly before his time. The -ta/ a- suffix is common in many North Indian tribal names such as in Mara a, Ara a, (K)ul ta (Witzel 1999).

first detailed grammar is in the 1930s¹³¹ and we do not have texts earlier than that. Nevertheless, this small language is of great linguistic interest as it has now been shown to be related to the northern Caucasus languages (Cherkes, Chechen, etc.) and Basque on the French/Spanish border, with which it shares a number of unusual characteristics (below). It seems to represent the Indian remnant of the movement out of S.W. Asia to Europe at c. 40 000 years ago. The ancestral language, Macro-Caucasian, can be reconstructed,¹³² though this is, again, pursued by very few linguists and remains controversial with many others.

Beyond the recent collections of oral literature, we also have some – unstudied – names in early manuscript colophons and inscriptions of the first millennium CE¹³³ (such as *pr ava* and a Sanskritized *puru a* as local self-designation *Burusho*).¹³⁴ There also are some translations of Sanskrit titles of Buddhist texts into Old Bur. and, parallel, into Tibetan. They have almost completely gone unstudied.¹³⁵ Unfortunately, these translations seem to be only general paraphrases of the Skt. titles,¹³⁶ and they need much more study. The same is true for early place names in the Northwest, and also for inscriptions along the Indus river, as well as in Buddhist manuscripts and modern geographical names.

However, the Bur. language is actually very old in this area. A few Bur. words occur already in the *gveda*, such as *kil la* ‘biestings’ – the first milk of a cow after giving birth. Much more is to be expected on closer study¹³⁷ when etymologizing the data of the existing dictionaries.¹³⁸

Though a small language in a marginal area, *Burushaski* is

131 Lorimer 1935–38

132 Bengtson 1990, 2001, 2003

133 O. von Hinüber 1980, 1989

134 Tibetan *Bru-ža*, and as I assume, even RV *m ja-* (in *maujavata*) and in Avestan *m ža*, as the name of a mountain and people in the area.

135 Poucha 1959

136 For those that I have studied; see the earlier note by Poucha 1959.

137 See an initial list in Witzel 1999; Tikkanen 1988 on the *Burushaski* (and other) substrates of the Northwest.

138 Such as Berger 1988, Tiffou 2004, Tiffou and Morin 1989 Grammatical description by Berger and Anderson (forthcoming).

of great linguistic and historical interest. For example, its nouns have four genders, one more than Indo-European: human and divine masculine, human and divine feminine, countable objects and uncountable ones (as in their putative relatives, the northern Caucasus languages). This division also influences the suffixes for the numbers: singular/plural/countable/indefinite. Nouns frequently must have a possessive prefix, especially for body parts or relatives. Similar to its putative relation to Basque and reflection in French, the numerals are based on a pattern of twenties (20, 40 as 2x20, etc.).

The verb has three basic stems: past, present and consecutive; the latter functions as a gerund (as in Skt., Dravidian), which makes this a pan-Indian feature. Curiously, a verb can have four prefixes and six suffixes, such as for the subject, object, person, for negative forms, etc. As in modern NIA languages, the verb has both nominative and ergative syntax. The latter stresses the performer of an action (agent) of a verb, and marks it, for example, in Nepali by the case ending *-le*, as in *mai-le garyo/gareko* 'It was done by me' = 'I have done' (with transitive verbs), as opposed to *ma gayo* 'I have gone' (with intransitive verbs). The intransitive verbs mark just the subject or object, while transitive verbs mark both the subject and the object separately.

KUSUNDA AND ANDAMANESE

Kusunda, too, has some very unusual features.¹³⁹ It has recently been connected with Andamanese and the New Guinea languages.¹⁴⁰

Its sound system is unusual as its pattern neither agrees with that of the surrounding Indo-Aryan/Nepali nor of that of the Tib.-Burmese languages. Consonants are much more varied and include a retroflex series, a pan-Indian feature (but also a retroflex *ç* as in Dardic). These sounds are, however, not phonemic but interchange with dentals, etc. There also is a uvular series (contrasting with *k*,

139 Watters 2005, whose description is based on 3 speakers. There are, however, as B.K. Rana informs me, some 20 speakers left, dispersed over Central Nepal, while the total population, according to the recent census is 156.

140 Whitehouse, Usher and Ruhlen 2004

g,) and a glottal stop.

There are the standard 5 vowels [a i u e o] plus [ɔ]. As in Kashmiri, there is some interchange between [i/e] and [u/o], as well as [ɔ/a]. The vowels can appear, unlike in most S. Asian languages, in diphthongs [ui, eo] etc., three sequential vowels [aio, iu] etc., and even as double diphthongs such as [iu a] in kiw atn ‘I pinched it’, [iu] in piw go ‘peel it’!

As in Burushaski, nouns can be prefixed with possessive pronouns, and they are sometimes lexicalized by their third person pronouns.

The verb has a system of tense-aspect-modality distinction. Pronominal prefixes are also necessary in the verb system, and are different for transitive and intransitive verbs; however, they are additionally suffixed as well. Except for the unrelated Burushaski, this is unlike any other S. Asian language.

While there are some gerund-like suffixes, they are more complex than in other S. Asian languages and do not necessarily imply subordination, but another kind of parataxis.

ANDAMANESE

The various Andamanese languages are better known as early British administrators took interest in the islanders and their languages. Recently, New Delhi linguists have taken up the slack, and have begun to describe the few remaining languages (Great Andamense, Önge, Jarawa), while Sentinel remains unknown. After a few, unsuccessful trials at contact, the latter island has been put under quarantine in 2004 – perhaps the only way to sustain this small group of people: it is well known that upon contact with modern civilizations, such isolated tribes quickly die out: only some 40 Great Andamanese and some 90 Önge are left, plus maybe 250–300 Jarawa. These have recently chosen to establish limited contact with the S. Asian settlers on the Great Andaman Island, and their language can now be studied in some depth.¹⁴¹

Like the other languages of the Indian subcontinent,¹⁴²

141 Abbi 2006, Sreenathan 2001

142 With the exception of Proto-Munda, which has however a * .

Andamanese languages¹⁴³ have a set of retroflex consonants, which strengthens (with Australian) the substrate theory (see above). Like Burushaski, they also have personal prefixes, especially for body parts and relatives. Like Dravidian they have an inclusive/exclusive pronoun. They also share the pan-Indian feature of gerunds.

All of the above underlines the fact that the current Dravidian-Indo-Aryan picture of Indian languages is myopic. In the distant past – as is still visible in a few fragments – the subcontinent had a large variety of languages and language families. They can tell us much about its prehistory – if only these ‘additional’ small tribal languages would be studied by more than by the handful of scholars per language family who are now active in the field.

II. EPIGENETICS: THE SOUTH ASIAN LINGUISTIC AREA

So far, we have discussed current languages, their ancestors and their linguistic families by making use of the trees of ‘family’ relationships, in other words, of a cladistic and phylogenetic arrangement, like Darwin’s scheme in *The origin of species*. There is more to it.

I have already referred to the ‘interactions’ between languages. As mentioned, languages and language families have interacted in South Asia for the past c. 3000 years of known linguistic history, and of course before that.

The first appearance of anatomically modern humans in S. Asia (straight out of Africa, via Arabia), dates back to around 75–65 kya. They now seem to be attested in Tamil Nadu at c. 74 kya.¹⁴⁴ This automatically means that the predecessors of all later Indian languages and their linguistic families have interacted ever since that early period.

We may call such ‘superficial’ mutual influences of neighbouring languages as well as, importantly, of earlier now lost languages, ‘epigenetic features’.

143 See now Abbi 2006 and cf. Usher 2006

144 Though only by stone artifacts, not by skeletons; they may belong to the previous *Homo erectus*; see however, Petraglia, Michael D. et al. 2007.

In recent times,¹⁴⁵ epigenetic influences have underlined the study of the inheritance of certain genetic materials, as opposed to Darwinian cladistics and family trees. These are non-genetic variations that individuals, human or not, acquire during their lifetime and unexpectedly pass on to their descendants by epigenetic inheritance. Obviously, this trend interferes with the purely genetic (chromosomal, autosomal) inheritance from both one's father and mother. However, the new data show that epigenetic inheritance is ubiquitous, from bacteria to plants and animals.¹⁴⁶ 'Incorporating epigenetic inheritance into evolutionary theory extends the scope of evolutionary thinking and leads to notions of heredity and evolution that incorporate development.'¹⁴⁷

This concept of the passing on of acquired features is close to the old one of Jean-Baptiste Lamarck (1744–1829). According to him (1809), an animal could acquire certain useful traits over several generations, such as the gradual lengthening of the neck in giraffes that allowed it to forage on higher trees. Inheritance according to Lamarck is not due to the pressure of selection and eventual disappearance of the lineages of individuals that did not inherit a specific beneficial mutation, but due to straight and immediate inheritance acquired by a parent that was passed down the line of children and grandchildren.¹⁴⁸

However, instead of throwing out the Darwinian baby with the Lamarckian epigenetic bathwater, it has to be observed that such changes only occur in very specific locations on the genes, and therefore the general phylogenetic tree, the stemma, remains unchanged up to that point. It also remains to be seen whether the new traits are persistent enough to be transmitted beyond a few generations. For our current undertaking, this development is exciting but not fatal.

145 Cf. <http://www.sciencedaily.com/releases/2009/05/090518111723.htm>

146 Epigenetic inheritance has been explained by four mechanisms so far, for example, DNA methylation: methyls attach to the DNA and cause genes to be active or inactive, without changing the inherited DNA. If external influences, for example, chemicals, change the current methyl configuration, the new trait can be inherited.

147 Jablonka et al. 2009

148 Cf. <http://www.sciencedaily.com/releases/2009/05/090518111723.htm>

If we then study similar effects in language, we can in fact detect both the equivalents of phylogeny and epigenetics in linguistics. The phylogenetic ‘family tree’ is obvious as even the simple comparison, shown earlier, of numbers in Hindi, Tamil and Munda indicates. This is not that, say Telugu or Tulu, suddenly pick up Hindi numbers,¹⁴⁹ nor everyday words or general features of grammar (endings, prefixes, infixes). Neither does Marathi use Kannada numbers, nor does Mundari use Oriya or Telugu or Kurukh (Drav.) ones. The same applies to Tibeto-Burmese, Burushaski, etc.

However, the successful (Darwinian style) phylogenetic reconstruction of language families (by ‘trees’, clades) is interfered by the separate wave-like spread of certain restricted features within a dialect area of one language and, importantly, across language boundaries, and even across those of language families.

This is actually easy to understand. If we think of modern English in Britain and the USA, etc. these varieties now have words such as jungle, bungalow, pajama, yoga, or the financial(!) guru or the (political) pandit. The latter two words indicate well the incipient changes in meaning of borrowed terms.

How did this influx of Indian words occur? Obviously, due to the Indian epigenetic influence on English, while British speakers of English were residing in India. As we know, the opposite, from English to local languages, has occurred in most modern Indian languages. One merely has to listen to a filmy dialogue in Bollywood Hindustani, with its occasional insertions of ‘thank you’, etc. by Hindi-only speakers.

What we have here is the superficial influence by an unrelated language on Hindustani, Tamil, etc. that like epigenetic traits is ‘inherited’ by the children, grandchildren etc. of those who first took over these new words. However, just as in epigenetics, such influences do not necessarily last forever. For example, my mother tongue, German, was full of French words between c. 1650 and 1900, but many, if not most of them, have since been lost and Germans have acquired new loans. The language now is full of English words – frequently for quite unnecessary concepts. The

149 Though we see that occurring, for example, in the Newari of Kathmandu, in marketplace situations, where numbers are taken from Nepali.

same is found in Japanese, where we have had a number of foreign words¹⁵⁰ (written in a special syllabary, Katakana), but where over the past 20 years, a host of new English words have appeared, like ‘«point, topic, news, member, maker (of cars, products)’ – for which older Japanese terms existed – (many of them often coined by using old Chinese loans, just as we do in English with Graeco-Roman words, as with auto-mobile instead of car).

In South Asia, this trend has been going on since the Persian conquests of the Panjab and Sindh in 530 BCE: as we have seen earlier, first Old Persian, then Greek, followed by Turkic, New Persian/Arabic, Portuguese and English words, and now sushi and kungfu. However, all of this remains only superficial.

There also is another, more serious, continuous influence, by local (Indian) languages, across language boundaries and families, on each other that has been going on ever since we have recorded Indian languages.¹⁵¹ Again, this is not surprising.

The matter of deep mutual influence was first studied in the Balkan area of SW Europe in the twentieth century, where a number of languages have heavily influenced each other. This concerns languages such as Rumanian, modern Greek, Albanian, Bulgarian, Macedonian, Serbian and the Gypsy language Romani. In other words, languages from five different sub-families of Indo-European that are not mutually understandable at all: Romance (Latin), Greek, Illyrian (Albanian), Slavic, and Indo-Aryan (Romani).

These people have been living together in the same villages or towns for many centuries and thus have acquired a working knowledge of their individual neighbours’ language(s). However, not just vocabulary but also syntactical structures have been borrowed across language families, such as the Rumanian postpositioned article –ul, which echoes the Bulgarian one in –yat: they are two quite different suffixes that carry the same meaning:

150 Such as mesu ‘knife’, trappu ‘staircase’, garasu ‘glass’ from Dutch, or kopfschmerz, neurose, sairu, ‘rope,’ hyutte ‘cabin’ from German.

151 Similarly, one can often translate sayings, collocations, concepts easily between western European languages (but less so with eastern European ones), and certainly not easily at all with other Indo-European ones in Iran and India or with those of other language families.

‘the’. This feature is unusual in the Balkans otherwise,¹⁵² but shared here by two neighbouring languages that belong to the Latin (Romance) and Slavic families. There are other features that cover all Balkan languages, such as the avoidance of the infinitive or future tense formation. We call this the Balkan linguistic area, or Sprachbund.

In the Balkan cases, the innovations have indeed been inherited by all succeeding generations of speakers, turning an epigenetic trait into a permanent feature of the local language in question. From a phylogenetic point of view, such a change may be called an unexplainable mutation.

In South Asia we have the same kind of clear influence, interference and convergence between Indo-Aryan, Drav., Munda and the rest, leading to the South Asian linguistic area, first mentioned by Emeneau in the 1960s.¹⁵³ As we have seen, such influences extend to syntax. Similar developments can be observed in South Asia.

Leaving aside the mutual borrowing of words across language boundaries, there now is the same word order in the Munda (Austro-Asiatic), Indo-Aryan, Dravidian, and Burushaski languages: subject–object–verb (SOV). This order is quite common in Eurasia, but not so in Kashmiri or Chinese (SVO), or in modern W. European languages which earlier had an SOV, nor in Semitic (VSO). Importantly, the Munda SOV word order differs substantially from the original Austroasiatic one,¹⁵⁴ indicating the influence of neighbouring languages (IA, Drav.?), on Munda through three thousand years of interaction.

There also are clear indications of the influence of neighbouring languages on the syntactic features, appearing already in our oldest available text, the *gveda*. The three most obvious ones were highlighted by the late F.B.J. Kuiper of Leiden in the 1960s.¹⁵⁵

152 Curiously it has been introduced, after the Old Norse period, into the Scandinavian languages, but how remains unclear, cf. William A. Craigie 2008 *The outlook in Philology*. *Transactions of the Philological Society* 43:12–27

153 Emeneau 1956; Kuiper 1967

154 P. Donegan and D. Stampe 2004, at: <http://www.ling.hawaii.edu/austroasiatic/AA/rhythm2004.pdf>.

155 Kuiper 1967 <http://www.springerlink.com/content/tl306hw646806112/>

1. *iti* as a quotation marker at the end of a direct speech, like Bengali ...*bole* (*boliy*) Nepali *bhanne...*, Marathi ... *mh un* ‘having said’, *asa* ‘thus’; Old Tamil (Puran *n ru*) *enru-*, more common: *ena*; modern Tamil *enru-*, *ani*, Kannada *endu*, Telugu *ani* ‘having said’. Cf. also the parallel occurrence in Japanese ... *to*, etc. This is also seen in Munda (Santali, Mundari: *mente*, *gamle* ‘having said’), but this probably is only a new formation under the influence of Indo-Aryan (and Drav.); notably, it is missing in the related Khasi language of the Shillong Hills.

2. *-tv / -ya*, two suffixes marking the so-called absolutes ‘having gone...’, etc.. They are seen in Skt. *gatv*, *sam-gam-ya*, which are echoed by two formations in Dravidian: Tamil (etc. South Drav. I) - *tu*, *-ntu*, *-ttu*; South Drav. II: **-cc-i*, etc.¹⁵⁶ However, such forms are also found in Burushaski, etc. (as well as in Indo-Aryan, in Nepali *-era*). The appearance in Bur. points to a possible early spread of this phenomenon in the Northwest. On the contrary, probably new formations occur in Munda: Santali *sen-kate* ‘having gone’, Mundari *hiju-akan-te* ‘having come’, Korku *hadir-en-ten* ‘having arrived’, as well as in Kharria *col-kon* ‘having gone’, Sora *jum-le* ‘having eaten’, Gorum *su u- u* ‘having fallen’. Again, the formation is not found in Khasi.

3. Retroflexes are consonant sounds that are produced with ‘bending the tongue backward’ as in *ṣ*, *ṣh*, *ṣr*. They are found in most Indian languages, but were originally lacking in Tib.-Burm. and in Munda (here, with exception of *ṣ*). Generally, they are supposed to have been taken over from Drav. (also by Kuiper), as if they always had been part and parcel of that language family. If common origin of Dravidian, Indo-European, Uralic, Altaic, etc. (as Nostratic) is accepted, they must be an innovation in Dravidian. The source may very well have been the very ancient Australian substrate in Dravidian, described earlier. Australian languages have them and sound, to the untrained ear, quite like Dravidian.

Further, it is remarkable that the heaviest concentration of retroflexes is found not in South India but in the Northwest, where the palatals *c*, *ch*, *j*, *jh* and even vowels(!) have a retroflex

variety. Vedic Skt. *yak a is Kalash j.a . or Kalasha p ~ < p r a-*, for the Fall festival. Retroflexion is found in all northwestern language families, be they Indo-Aryan, Nuristani, Iranian (Pashto, Khotanese Saka) or Burushaski. In other words, retroflexion has impacted everyone coming into the subcontinent and some closely neighbouring areas, such as Khotanese Saka (in SW Xinjiang) or the eastern (not the western) dialects of Baluchi, a late immigrant language from western Iran, arriving only after 1000 CE.

In sum, retroflexion is a regional NW and only then a pan-Indian feature. Dravidian may have acquired it upon arrival in the subcontinent; when is another question.¹⁵⁷

Other markers, such as those highlighted by the late M. Emeneau¹⁵⁸ include the ubiquitous ‘expressives’ for onomatopoeics, peculiar actions, etc. In Indo-European we mostly find substitutes by regular verbs such as murmur, or the occasional splash-splash. But they are extremely common in much of Asia (especially in Japanese), though they are rarely listed in our dictionaries.¹⁵⁹ My impression of their types in South Asia is that Indo-European, Drav. and Munda are different in this respect. Indo-European now has the *bal-bal::kara-kara::ro i- o i* types. The first two are found in Vedic texts.¹⁶⁰ The *bal-bal* type seems to be of Indo-European vintage, and such collocations have remained unchanged through history, while other words have changed:¹⁶¹ *bal-bal* ‘sound of emptying a water vessel’ is still *bal-bal* today in Hindi, etc. while it should have become something like *balla* or rather *b l*. The second type is typical for Dravidian:¹⁶² Tamil *kara-hara*, while the third,

157 Their linguistic ancestors, may have come with the Borean speakers (see above), straight out of Africa, via Arabia. If Dravidian indeed formed part of the Nostratic super-family, it will have split off from it somewhere in the Greater Near East, along with Uralic, with which is often linked.

158 Emeneau 1956; B. Krishnamurti 2003: 39

159 Emeneau 1969

160 Hoffmann, Karl. *Wiederholende Onomatopoeitika im Altindischen. Aufsätze zur Indoiranistik*, Wiesbaden 1987/1991 [=Indogermanische Forschungen 1952, 60, 254–264]

161 The same is true of Japan: *peko-peko* ‘sound of a hungry stomach’, which should have become **heko-beko*, due to the development of modern Japanese, but has stayed at the Old Jpn. level of development.

162 Also in Drav., see Krishnamurti 2003: 131, 136 sq. with examples such as

the ro i- o i type, is typical for Munda: Mundari ma i-u i ‘food and the like’, jaldi- paldi ‘quickly’; Kharia: o iga ‘eat up’!, Sora: du du ‘thud, thump’, mand n tad n ‘plates’ (mand n dish), br i b r ‘to dazzle’.¹⁶³

What we see in all these features is a certain amount of convergence between neighbouring languages and language families, due to long term adjacent habitation, resultant social interaction, and bi- or trilingualism. In other words, the successful (Darwinian style) phylogenetic reconstruction of language families receives interference by the separate wave-like spread¹⁶⁴ of certain features across linguistic boundaries. The S. Asian language families have increasingly become more similar to each other over the past few millennia. Their shared, common traits have been inherited since the original contact, even if they are expressed by the individual means of the language in question.

It seems, in the footsteps of Darwin, that these changes have been mutually beneficial. Due to common features, one can shift more easily from one language’s system into that of the other, without having to reconfigure or ‘reorder one’s brain’, as one has to do when one shifts from Hindi or Tamil to English and vice versa. One can easily see this when one tries to translate a complex sentence with relative clauses or indirect speech back and forth.

This means quicker and easier communication with neighbours, people from other villages, in the marketplace, and so on. Or even in the same household, like many South Asians, I had 5 languages spoken in my house at Kathmandu back in the 1970s.¹⁶⁵ The advantage is thus one of serendipity, or ‘laziness’: one can stay in one ‘mold’ and continue speaking – as we know in S. Asia – while changing the language even within one and the same

*gu V-gu V ‘gurgling sound’ > Kannada gud -gu u, Telugu gu a-gu a, etc.

163 Anderson 2008

164 Something actually studied, for Indo-European dialects, since the 1870s, first by Johannes Schmidt (his Wellentheorie = ‘wave theory’ 1872, as opposed to August Schleicher’s Stammbaumtheorie = phylogeny, 1853/1860)

165 That is, incidentally, why speakers of Japanese learn Hindi, Nepali or Tibetan very quickly as these unrelated languages have a grammatical (syntactical) structure similar to Japanese.

sentence.¹⁶⁶

What we witness here is a gradual convergent development of South Asian languages, based on the gradual shift from their respective Indo-European, Drav., Austro-As. parent languages to their various modern forms, belonging to three major family trees, next to the Tib.-Burm., Macro-Caucasian (Burushaski), Indo-Pacific (Andamanese, Kusunda) ones on the rims of the subcontinent. The result was the South Asian Sprachbund.¹⁶⁷

In sum: a combination of both the phylogenetic and the epigenetic features lead to the emergence of the current South Asian linguistic area (sprachbund). This development has made the structure of Indo-Aryan, Dravidian or Munda increasingly similar to each other, but it could not eliminate most of their individual characteristics.

MISINTERPRETATIONS AND OUTLOOK

Nowadays, the South Asian sprachbund feature has been misinterpreted by some who are not necessarily specialists in linguistics. There is a pervasive feeling that every native speaker ‘is a language specialist’. However, you still cannot understand Tamil if you are a Panjabi or Bengali speaker, and a Munda speaker cannot understand Hindi or Malayalam in the same rough way that a Hindi speaker may understand Marathi or Nepali. This is abundantly clear from the list of numbers from 1 to 10 given

166 To give an example from personal experience: in my own multilingual family (Japanese–German–English), we do not only switch between languages but have even created our own Creole, based on pidginized German, with heavy Japanese grammatical influence and a little English adstrate, especially in vocabulary: open syllables (ending in vowels, which mean, for example, feminine adjectives) are preferred, the article is dropped, only two tenses (present/past) are used, etc.

167 If all these languages had been isolated from each other, we would now have results such as that seen in the isolated Icelandic compared to Scandinavian, or Maori compared to Hawaiian, Lithuanian compared to other European languages, Toda compared to Tamil, Kalasha and Khowar compared to the other Dardic languages like Kashmiri and the neighbouring Indo-Aryan language, Panjabi. However, most S. Asian languages have been close neighbours of each other, often in the same villages, like the isolated Nahali and the Munda language Korku on the Tapti river. They thus have heavily influenced each other.

earlier, or basic words like ‘to go’ or ‘hand’ that remain separate and mutually unintelligible even today.¹⁶⁸

Engl.	hand	eye
Hindi	<i>hath</i>	<i>āñkh</i>
Tamil	<i>kai</i>	<i>kaṇ</i>
Korku &	<i>tī</i>	<i>me’d</i>
Mundar	<i>ī</i>	

In sum, even some 3000–5000 years of geographical and physical neighbourhood of the South Asian language families involved could not erase the inherited ‘genetic’ characteristics of the various Indian languages. Therefore, the talk of being ‘pro-Aryan’ or ‘pro-Dravidian’ or anti-x, -y, -z is just political talk. The language, religion, rituals of the earliest texts, or the cline of genetic features from South to North indicate that South Asia is a region well defined vis-à-vis the rest of Asia, but is one with many remaining, clear internal differences, a fact that should not really perturb anybody today.¹⁶⁹

Instead, as scholars, we should rather investigate the overlaps of our respective fields, if any, and see where a particular language and its historical and contemporary spread has a clear correlation

168 This is not unlike the position of Finnish, Hungarian or Basque in Europe: it is immediately visible that they belong to different families. But they are fairly easily translatable into/from other W. European languages due to the same, several thousands of years old cultural background. So is Telugu/Hindi.

169 To compare the European situation again: speakers of Finnish, Hungarian and Estonian, which do not belong to the Indo-European but to the Uralic language family, are just as good Europeans as the rest as they have shared a common culture for one or more millennia. This also includes the non-Indo-European speakers of Basque, of Maltese (speaking a variety of Arabic), the Muslim Albanians and Bosnians, the Caucasian-speaking Georgians, and the Indo-European Armenians. Due to complex historical and cultural reasons, it is more difficult to be accepted as a speaker of an Altaic language, such as Turkish with a Near Eastern Islamic culture (but contrast the Slavic-speaking Muslim Bosnians or ‘Illyrian’-speaking Albanians), or the Mongolian-speaking Buddhist Kalmyks in S. Russia. Both only marginally share pan-European culture, and that is crucially important, irrespective of religion and language. In the same sense, all of South Asia is one region that includes Pakistan, India, Nepal, Bangladesh, Bhutan, Sri Lanka and the Maldives, but significantly not Afghanistan, Burma or other neighbours.

in archaeology, human paleontology, population genetics, linguistics, texts, etc. It is this that we have been studying at Harvard Round Tables dedicated to South and Central Asia since 1999, and we have reached some consensus.

However, this kind of interdisciplinary conversation and the ensuing comparison of data of the natural and social sciences and of linguistics, religion, ritual, mythology still is weak in the subcontinent. These fields of the humanities should be developed just as archaeology and human population genetics have been in South Asia over the past decades. All of them contribute in telling us a facet of where we have come from, what we have retained from earlier periods, and thus, who we are today.

For, we are complex beings and, in order to understand ourselves, we must achieve this kind of scientific cooperation. Again, the present conference was a good opportunity to underline this principle. Normally, we all are engulfed in the detailed, specialist work of our professions, and that is certainly necessary. But we should also look beyond these narrow confines. If we can establish multidisciplinary and multinational cooperation between our various fields, wherever possible, that would be a most beneficial outcome of this Jubilee meeting.

REFERENCES

Abbi, Anvita. *Endangered Languages of the Andaman Islands*. Munich: Lincom, 2006

Anderson, Gregory D. S. "Recent Advances in Proto-Munda and Proto-Austroasiatic Reconstruction." Paper presented at the 3rd Harvard Round Table on the Prehistory of Central and South Asia, Cambridge, MA. (May 2001)

---, *A New Classification of South Munda: Evidence from Comparative Verb Morphology*. Indian Linguistics Poona: Linguistic Society of India 62, 2001: 21–36

---, *The Munda verb: typological perspectives*. Berlin, New York: Mouton de Gruyter, 2007

---, *The Munda languages* / edited by Gregory S. Anderson. London ; New York : Routledge, 2008

---, *Burushaski*. In *Language Islands: Isolates and Microfamilies of Eurasia*, ed. by D.A. Abondolo. London: Curzon Press (forthcoming)

Arensburg B., et al. *A Middle Palaeolithic human hyoid bone*. *Nature*.

1989, Apr 27;338, (6218), 758-60

Bajracharya, Dhanavajra, *The Goplarjava val.* [edited and translated by] Dhanavajra Vajr c rya and Kamal P. Malla. Wiesbaden, GmbH: Franz Steiner Verlag, 1985

Beekes, Robert S.P. *Comparative Indo-European linguistics: an introduction*, Amsterdam, Philadelphia: John Benjamins, 1995

Bender, M. Lionel. *The Nilo-Saharan Languages: A Comparative Essay*. München: Lincom Europa, 1996

Benedict, Paul K. *Sino-Tibetan, a conspectus* [by] Paul K. Benedict. Contributing editor: James A. Matisoff. Cambridge, University Press, 1972.

---, *Austro-Tai: Language and Culture, with a Glossary of Roots*. New Haven: HRAF Press 1975

---, "Austro-Thai and Austroasiatic." In: *Austroasiatic Studies, Part I*, ed. Philip N. Jenner, Laurence C. Thompson, Stanley Starosta, 1-36. Honolulu: University of Hawai'i Press, 1976

---, *Japanese/Austro-Tai* / Paul K. Benedict. Ann Arbor: Karoma, 1990

Bengtson, John D. "An End to Splendid Isolation: The Macro-Caucasian Phylum." *Mother Tongue* [Newsletter] 10 (April 1990) ?? [cf. [http://www.nostratic.ru/books/\(220\)bengtson%20-%20macro-caucasic.pdf](http://www.nostratic.ru/books/(220)bengtson%20-%20macro-caucasic.pdf)]

---, "Genetic and Cultural Linguistic Links between Burushaski and the Caucasian Languages and Basque". Presentation at the 3rd Harvard Round Table on the Ethnogenesis of South and Central Asia, 2001.

[http://www.nostratic.ru/books/\(299\)bengtson%20-%20BurshRoundtable.pdf](http://www.nostratic.ru/books/(299)bengtson%20-%20BurshRoundtable.pdf)

---, "A Multilateral Look at Greater Austric." *Mother Tongue* XI, (2006) 219-258

Bengtson, John and Merritt Ruhlen. *Global etymologies*. 1994 <http://www.merrittruhlen.com/files/Global.pdf>

Berger, Hermann *Die Burushaski-Sprache von Hunza und Nagar*. Vols. I-III. Wiesbaden: Harrassowitz 1988

Bickerton, D. *Language and Species*. Univ. of Chicago Press 1990

Black, Paul. *Outside relationships of Australian languages*. *Mother Tongue* XI, 2006, 259-261

Blažek, Vaclav. "Was there an Australian Substratum in Dravidian?" *Mother Tongue* XII (2006): 275-294

Louis-Jean Boe et al. *The potential Neandertal vowel space was as large as that of modern humans*. *Journal of Phonetics* 30, 2002, 465-484 doi:10.1006/jpho.2002.0170. <http://www.ideallibrary.com>

Bonfante, Giuliano & Larissa Bonfante. *The Etruscan Language: an Introduction*. Manchester: University of Manchester Press 2002

Burrow, Thomas and M. B. Emeneau, A Dravidian Etymological Dictionary. Oxford: Clarendon Press 1960

Bryant, Edwin & Laurie Patton. The Indo-Aryan controversy : evidence and inference in Indian history / edited by Edwin F. Bryant and Laurie L. Patton. London & New York : Routledge, 2005

Bub, Daniel. Reflections on language evolution and the brain. Review of Phillip Lieberman's *Toward an Evolutionary Biology of Language*. *Cortex* 44, 2008, 206-217

Callaway, Ewen. Neanderthals speak out after 30,000 years. *New Scientist*, 15:00 15 April 2008.

<http://www.newscientist.com/article/dn13672-neanderthals-speak-out-after-30000-years.html?DCMP=ILC-hmts&nsref=spectr10_head_Neanderthal%20talk>

Capell, Arthur. "History of Australian Languages." In *Australian linguistic studies*, ed. Stephen A. Wurm, 419-619. Canberra: Dept. of Linguistics, Research School of Pacific Studies, Australian National University 1979

Cardona, George and Dhanesh Jain (eds.). *The Indo-Aryan languages*. London; New York: Routledge 2003

Chatterji, Suniti Kumar. *The origin and development of the Bengali language*. With a foreword by Sir George Abraham Grierson. Calcutta: Calcutta University Press 1926

---, *Kir ta-jana-k ti = The Indo-Mongoloids, their contribution to the history and culture of India*. Rev. 2d ed. Calcutta: Asiatic Society 1974

Clendon, Mark. "Reassessing Australia's Linguistic Prehistory." *Current Anthropology* 47, 2006, 39-61

Culotta, Elizabeth. "Ancient DNA Reveals Neandertals With Red Hair, Fair Complexion." *Science* 318, Oct. 25 2007, 546-547

Décsey, Gyula. "The Uralic Protolanguage : a Comprehensive Reconstruction. Etude descriptive générale du proto-ouralien et glossaires proto-ouralien-anglais." *Ural-altaische Jahrbücher* 62 (1990) 9-147

Diffloth, Gérard. *An appraisal of Benedict's views on Austroasiatic and Austro-Thai relations*. Kyoto: Center for Southeast Asian Studies, Kyoto University, 1976 [Discussion paper (Kyoto Daigaku. Tonan Ajia Kenkyu Senta) no. 82]

---, *The Lexical Evidence for Austric, So Far*. *Oceanic Linguistics*, 33, 1994, 309-321

---, "Tentative Calibration of time depths of Austroasiatic branches." Presentation at the Colloque "Perspectives sur la Phylogénie des Langues d'Asie Orientales" at Périgueux, France, 30 August 2001

---, *The contribution of linguistic palaeontology to the homeland of Austro-asiatic*. In: Sagart, Laurent, Roger Blench and Alicia Sanchez-

Mazas (eds.). *The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics*. RoutledgeCurzon 2005: 79–82

Dillon, Anthony V.N., Jerold A. Edmondson, Yongxian Luo. *The Tai-Kadai Languages*. London/New York 2008

Dixon, Robert M.W. *The Languages of Australia*. New York: Cambridge University Press, 1980

--- *Australian Languages. Their Nature and Development*. Cambridge: Cambridge University Press 2002

Donegan, Patricia and David Stampe. Rhythm and the synthetic drift of Munda. In: Rajendra Singh, ed., *The Yearbook of South Asian Languages and Linguistics*. Berlin and New York: Mouton de Gruyter: 2004, 3-36, <http://www.ling.hawaii.edu/austroasiatic/AA/rhythm2004.pdf>

Ehret, Christopher. *Reconstructing Proto-Afroasiatic (Proto-Afrasian). Vowels, Tone, Consonants and Vocabulary*. Berkeley: University of California Press, 1995

---, *A Historical-Comparative Reconstruction of Nilo-Saharan*. Köln: R. Köppe-Verlag 2001

Emeneau, M. B. India as a linguistic area. *Language* 32, 1956, 3-16

---, Onomatopoeics in the Indian linguistic area. *Language* 45, 1969, 274-299

---, and Th. Burrow, *Dravidian Borrowings from Indo-Aryan*. Berkeley : University of California Press 1962

Enard, W. et al. A humanized version of *FOXP2* affects cortico-basal ganglia circuits in mice. *Cell* 137, 2009, 961-71. doi:10.1016/j.cell.2009.03.041.

Enrico, John. Toward Proto-Na-Dene. *Anthropological Linguistics* 46, 2004, 229 – 302

Fisher, Simon. Tangled webs: Tracing the connections between genes and cognition. *Cognition*. 101, 2006 Sep, 270-97.. doi:10.1016/j.cognition.2006.04.004; <http://www.ncbi.nlm.nih.gov/pubmed/16764847>

Foley, William A.. *The Papuan Languages of New Guinea*. Cambridge: Cambridge University Press, 1986.

Fortescue, Michael. *Comparative Chukotko-Kamchatkan Dictionary*. Berlin: Mouton de Gruyter 2005

Greenberg, Joseph Harold. *Language in the Americas*. Stanford: Stanford University Press, 1987

Harrod, James. “Periods of Globalization over ‘The Southern Route’ in Human Evolution (Africa, Southwest Asia, Southern Asia, Southeast Asia, and Sahul and the Far East). A Meta-Review of Archaeology and Evidence for Symbolic Behavior.” *Mother Tongue XI* (2006): 23-84

---, *Researching the Origins of Art, Religion and Mind*. Available: <http://www.originsnet.org/glossmeth.html>.

Heine, Bernd. *African languages: an introduction*; edited by Bernd Heine and Derek Nurse. Cambridge: Cambridge University Press 2000

Hewitt, B. G. *Georgian: a structural reference grammar*. Amsterdam: John Benjamins 1995

Hill, Nathan Wayne. *An overview of the Tibetan verbal system*. Harvard PhD Dissertation, 2003

Hinüber, O. v. *Die Kolophone der Gilgit-Handschriften*, *Studien zur Indologie und Iranistik* 5/6, 1980, 49-82

---, *Brhm inscriptions on the history and culture of the upper Indus valley*. In: K. Jettmar et al., *Rock Inscriptions in the Indus Valley. Antiquities of Northern Pakistan. Reports and Studies*, vol. 1. Mainz: Philipp von Zabern 1989: 41-72

Hoffmann, Karl. *Die alt-indoarischen Wörter mit . Ph. Diss.* München 1941

---, *Aufsätze zur Indoiranistik*. (ed. J. Narten, vols.1-2) Wiesbaden. 1975-76

---, *Aufsätze zur Indoiranistik*. (ed. S. Glauch, R. Plath, S. Ziegler, vol. 3). Wiesbaden 1992

Hualde, José Ignacio & Ortiz de Urbina, Jon (eds.). *A Grammar of Basque*. Berlin: Mouton de Gruyter, 2003

Jablonka, Eva et al. *Transgenerational Epigenetic Inheritance: Prevalence, Mechanisms, and Implications for the Study of Heredity and Evolution*. *The Quarterly Review of Biology* 84, 2009, 131 DOI: 10.1086/598822

James, Hannah V.A. "The emergence of modern behavior in South Asia: A review of the current evidence and discussion of its possible implications." In *The Evolution and History of Human Populations in South Asia. Inter-disciplinary Studies in Archaeology, Biological Anthropology, Linguistics and Genetics*, ed. Michael D. Petraglia and Bridget Allchin, 201-227. Dordrecht: Springer 2007

Junker, Thomas. *Die Evolution des Menschen*. München: Beck 2006.

Kölver, Bernhard and Hemrj kya. *r Rudravar a Mah vih ra (P tan, Nepal) Documents from the Rudravar a-Mah vih ra, P an*. Sankt Augustin: VGH Wissenschaftsverlag 1985

Koppers, Wilhelm. *Die Bhil in Zentralindien*. Horn: F. Berger 1948

Krause, Johannes et al. "The Derived FOXP Variant of Modern Humans was shared with Neanderthals." *Current Biology* 17 (21), 2007, 1908-12

Krishnamurti, Bhadriraju. *Comparative Dravidian linguistics : current perspectives*. With a foreword by Murray B. Emeneau. Oxford ;

New York: Oxford University Press 2001

---, The Dravidian languages. Cambridge ; New York: Cambridge University Press 2003

Kuiper, F.B. J. The genesis of a linguistic area. *Indo-Iranian Journal* 10, 1967, 81-102, <http://www.springerlink.com/content/tl306hw646806112/fulltext.pdf>

---, Nahali, a comparative study. Amsterdam: Noord-Hollandse Uitgevers Maatschappij 1962

Lieberman, Philip. *Toward an evolutionary biology of language*. Harvard University Press, 2006

Lieberman & McCarthy, 2007:

<http://penn.museum/documents/publications/expedition/PDFs/49-2/Lieberman.pdf>; *Cortex* 44, 2008, 218-226

Lorimer, D. L. R. *The Burushaski language*. Oslo: H. Aschehoug; Cambridge, Mass.: Harvard University Press, 1935-38

Mahadevan, Irvatham. *Early Tamil epigraphy from the earliest times to the sixth century A.D.* Chennai, India: Cre-A; Cambridge, MA : Dept. of Sanskrit and Indian Studies, Harvard University. Distributed by Harvard University Press, 2003

Masica, Colin P. *Aryan and Non-Aryan Elements in North Indian Agriculture*. In: Madhav M. Deshpande and Peter Edwin Hook (eds.), *Aryan and Non-Aryan in India*. Ann Arbor: Center for South and Southeast Asian Studies 1979: 55-151

Matisoff, J. *Handbook of Proto-Tibeto-Burman: system and philosophy of Sino-Tibetan reconstruction*. Berkeley : University of California Press 2003

Mayrhofer, Manfred. *Etymologisches Wörterbuch des Altindiarischen*. Heidelberg: C. Winter 1986-2001

MacAndrew, Alec. "FOXP2 and the Evolution of Language." *Nature* 418, 2002, : 869-872

Misra, Satya Swarup. *The Aryan Problem. A linguistic approach*. New Delhi: Munshiram Manoharlal 1992

Nikolayev, S., and S. Starostin. *North Caucasian Etymological Dictionary*. Moscow: Asterisk Press 1994

<http://starling.rinet.ru/>

Noonan James P. et al, "Sequencing and Analysis of Neanderthal Genomic DNA." *Science* 314, Issue 5802, November 17, 2006, 1113 - 1118. <http://www.sciencemag.org/cgi/content/abstract/314/5802/1113?etoc>

Nurse, Derek and Gérard Philippson. *The Bantu Languages*. London, New York: Routledge, 2003

Ouattaraa, Karim, et al. Campbell's monkeys concatenate vocalizations into context-specific call sequences.

Proceedings of the National Academy of Sciences, 116, December 22, 2009, no. 51: 22026-22031. <http://www.pnas.org/content/106/51/22026.abstract?etoc>

Pawley, Andrew and Malcolm Ross. "The prehistory of Oceanic languages: A current view." In *The Austronesians: Historical and Comparative Perspectives*, ed. Peter S. Bellwood, James J. Fox, D. T. Tryon. Canberra: Australian National University 1995: 39-74

Petraglia, Michael D. and Bridget Allchin, eds. *The Evolution and History of Human Populations in South Asia. Inter-disciplinary Studies in Archaeology, Biological Anthropology, Linguistics and Genetics*. Dordrecht: Springer 2007

Petraglia, Michael D. et al. Middle Paleolithic Assemblages from the Indian Subcontinent Before and After the Toba Super-Eruption. *Science* 317, 2007, 114-116

Pinnow, Heinz-Jürgen. *Versuch einer historischen Lautlehre der Kharia-Sprache*. Wiesbaden, O. Harrassowitz 1959

Poucha, Pavel. "Bruzha -- Burushaski?" *Central Asiatic Journal* 5, 1959, 295-300

Powell, Adam et al. Late Pleistocene Demography and the Appearance of Modern Human Behavior. *Science* 324, no. 5932, 5 June 2009, 1298 - 1301 DOI: 10.1126/science.1170165

<http://www.sciencemag.org/cgi/content/abstract/324/5932/1298>. [Overview story in *Science Daily*:

<http://www.sciencedaily.com/releases/2009/06/090604144324.htm>]

Rana, Bhai K. *New Materials on the Kusunda language*. Paper presented at the Fourth Harvard Roundtable on the Ethnogenesis of South and Central Asia, Harvard University, May 11-13, 2002 <http://www.fas.harvard.edu/%7Esanskrit/RoundTableSchedule.html>

Robbeets, Martine Irma. *Is Japanese related to Korean, Tungusic, Mongolic and Turkic?* Wiesbaden: Harrassowitz 2005

Ruhlen, Merritt. *On the Origin of Languages, Studies in Linguistic Taxonomy*. Stanford: Stanford University Press 1994a

---, *The Origin of Language : tracing the evolution of the mother tongue*. New York: Wiley 1994b

<http://www.merrittruhlen.com/files/Global.pdf>

Sagart, Laurent (ed.) *The Peopling of East Asia: putting together archaeology, linguistics, and genetics*. London: Routledge 2005

Schmid, L. Report on a survey of Dardic languages of Kashmir. *Indian Linguistics* 42, 1981, 17-21

Schmitz, Ralph W. "Interdisziplinäre Untersuchungen an den Neufunden aus dem Neandertal. Johann Carl Fuhlrott (1803 - 1877) gewidmet." *Mitteilungen der Gesellschaft für Urgeschichte* 12 (2003) 25. <http://www.urgeschichte.uni-tuebingen.de/fileadmin/downloads/>

GfU/2003/gfu12_3.pdf

Schrenk, Friedemann, Stephanie Müller. Die Neanderthaler. Unter Mitarbeit von Christine Hemm. München: Beck, 2005. [English trans.: The Neanderthals, translated by Phyllis G. Gestice. London: Routledge, 2009]

Sen, Sukumar. An etymological dictionary of Bengali, c. 1000-1800 A.D. Calcutta, Eastern Publishers 1971

Southworth, Franklin C. Linguistic archaeology of South Asia. London/New York: RoutledgeCurzon 2005

Spalding, Alex. Inuktitut: a Grammar of North Baffin Dialects. Winnipeg, Canada: Wuerz 1992.

---, Inuktitut: A multi-dialectal outline dictionary (with an Aivilingmiutaq base). Iqaluit, NT: Nunavut Arctic College 1998

Sreenathan, M. The Jarawas : language and culture. Calcutta: Anthropological Survey of India, Govt. of India 2001

Solymosi, Maria Agnes et al: A Window on the Genetics of Human Speech: the *Foxp2* Gene.

Journal of the Association for the Study of Language in Prehistory, Issue XII, 2007, 135-142 www.mtakpa.hu/kpa/download/1209118.pdf

Stampe, David. [various papers on Austro-Asiatic]

<http://www.ling.hawaii.edu/austroasiatic/>

Sverdrup, Harald. A study of the Tartessian Script and Language. In: Shervoroshkin, Vitaly and Paul Sidwell. Languages and their Speakers in Ancient Eurasia. Canberra: Association for the History of Language 2002: 115-141

T van yan, N̄ namuttanetal. Centamilccorpirappiyal rakaramutali. Comprehensive etymological dictionary of the Tamil language. Cennai: Centamilc Corpirappiyal Akara Mutalit Tit t a Iyakkaka Vel y t u, 1985-

Thangaraj, Kumarasamy. "Genetic Affinities of the Andaman Islanders, a Vanishing Human Population." Current Biology 13, January 21, 2003: 86–93

---, et al., Reconstructing the origin of the Andamanese islanders. Science 308, 2005, 996

Thurgood, Graham and Randy J. LaPolla. The Sino-Tibetan Languages. London/New York: Routledge 2003

Tiffou, Étienne (ed.). Bourouchaskiana : actes du colloque sur le bourouchaski organisé à l'occasion du XXXVIème Congrès international sur les études asiatiques et nord-africaines (Montréal, 27 août-2 septembre 2002). Louvain-la-Neuve: Peeters 2004

--- and Morin, Yves Charles. Dictionnaire complémentaire du Bourouchaski du Yasin. Paris : Peeters/SELAF 1989

Tikkanen, Bertil. On Burushaski and other ancient substrata in northwest South Asia. *Studia Orientalia* 64, 1988, 303–325

Tishkoff, Sarah A. "History of Click-Speaking Populations of Africa Inferred from mtDNA and Y Chromosome Genetic Variation." *Molecular Biology and Evolution* 10 (2007 Oct. 24): 2180–95

Trinkhaus, Eric. "Human Evolution: Neanderthal GEME Speaks out." *Current Biology* 17 (21), 2007, R917-R919

Tulpule, Shankar Gopal and Anne Feldhaus. A dictionary of Old Marathi. Assisted by M.P. Pethe ; lexicographical consultant: Jayashree Gune. Oxford ; New York: Oxford University Press 2000

Turner, R. L. A comparative Dictionary of the Indo-Aryan Languages. London 1966 (CDIAL)

van Driem, George. "Tibeto-Burman vs Indo-Chinese: implications for population genetics, archeologists, and prehistorians." In *The peopling of East Asia: putting together archaeology, linguistics, and genetics*, ed. Laurent Sagart, et al. London: Routledge 2005: 81-106

---, "The prehistory of Tibeto-Burman and Austrasiatic in the light of emergent population genetic studies." *Mother Tongue XI* (2006) 160-211

---, "Austroasiatic phylogeny and the Austroasiatic homeland in light of recent population genetic studies". *Mon-Khmer studies: a journal of Southeast Asian languages and cultures* 37, 2007, 1–14

Usher, Timothy. Great Andamanese reconstruction underway. A condensed handout with tentative remarks on Papuan and Australian vis-à-vis external language families. *Mother Tongue XI*, 2006, 295-298

Vajda, Edward J. Ket. (2004) Munich: Lincom Europa 2004

Vovin, Alexander. A reconstruction of Proto-Ainu. Leiden ; New York: E.J. Brill 1993

---, "Introduction." In *Perspectives on the origins of the Japanese language*, ed. Toshiki Osada and Alexander Vovin, with the assistance of Kerri Russel.. Kyoto: International Center for Japanese Studies 2003: 421-428

Watters, David E. Notes on Kusunda grammar (a language isolate of Nepal), with the participation of Yogendra P. Yadava, Madhav P. Pokharel, Balaram Prasain. Kathmandu: National Foundation for the Development of Indigenous Nationalities 2005

[Watters' Kusunda dictionary at the SARVA website: http://www.aa.tufs.ac.jp/sarva/materials_frame.html]

Wells, Spencer. *The journey of man : a genetic odyssey*. Photographs by Mark Read. Princeton, N.J.: Princeton University Press 2002

Wells, Spencer et al: The Eurasian Heartland: A continental perspective of Y-chromosome diversity. *Proc. Natl. Acad. Sci. USA* 98, 2001, 10244- 9.

Whitehouse, Paul. The "Lost" Paper: A Belated Conference Postscript. *Mother Tongue* XI, 2006, 262-274

Whitehouse, Paul, Timothy Usher, Merritt. Ruhlen. "Kusunda: An Indo-Pacific language in Nepal." *Proceedings of the National Academy of Sciences* 101, 15, April 13, 2004, 5692–5695. www.pnas.org/cgi_doi_10.1073_pnas.0400233101

<http://www.pnas.org/cgi/reprint/101/15/5692.pdf>

Williamson, Kay & Roger Blench. "Niger-Congo'." In *African Languages - An Introduction*, ed. Bernd Heine, Bernd and Derek Nurse, 11-42. Cambridge: Cambridge University Press 2000

Witzel, Michael. On the location of the Licchavi Capital of Nepal. *Festschrift für P. Thieme (= Studien zur Indologie und Iranistik 5/6)*, 1980, 311-337

---, "Early Sources for South Asian Substrate Languages." *Mother Tongue*. Extra number. 1999 [cf. EJVS version, <http://www.ejvs.laurasianacademy.com/issues.html>]

---, "Comparison and Reconstruction: Language and Mythology." *Mother Tongue* VI, 2001, 45- 62

---, *Early Loan Words in Western Central Asia: Indicators of Substrate Populations, Migrations, and Trade Relations. Linguistic Evidence for Cultural Exchange in Prehistoric Western Central Asia*. Philadelphia: Sino-Platonic Papers 129, 2003; partial pdf: http://www.people.fas.harvard.edu/%7Ewitzel/Sinttashta_Qu.pdf

---, *The Origins of the World's Mythologies*. Oxford University Press. In Press, Fall 2010

Wurm, Stephen A. *Languages of Australia and Tasmania*. The Hague: Mouton 1972

---, *New Guinea Area Languages and Language Study, Volume 1: Papuan Languages and the New Guinea Linguistic Scene*. Pacific Linguistics, Research School of Pacific and Asian Studies, Australian National University. Canberra: Australian National University, 1977.

Zvelebil, Kamil. The descent of the Dravidians. *International Journal of Dravidian Linguistics* 1-2, 1972, 56-63

---, *Dravidian Linguistics: an Introduction*. Pondicherry: Pondicherry Institute of Linguistics and Culture 1990