

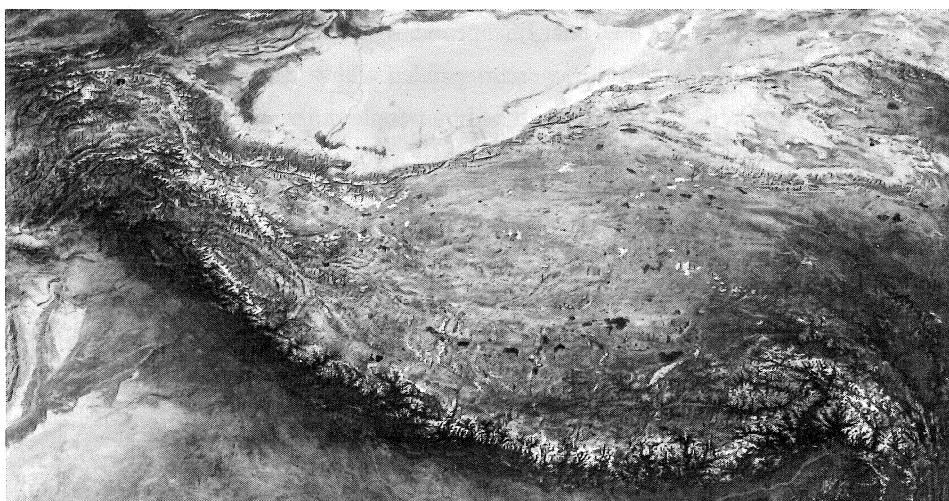
George van Driem

Professor of Linguistics at Leiden University in the Netherlands;

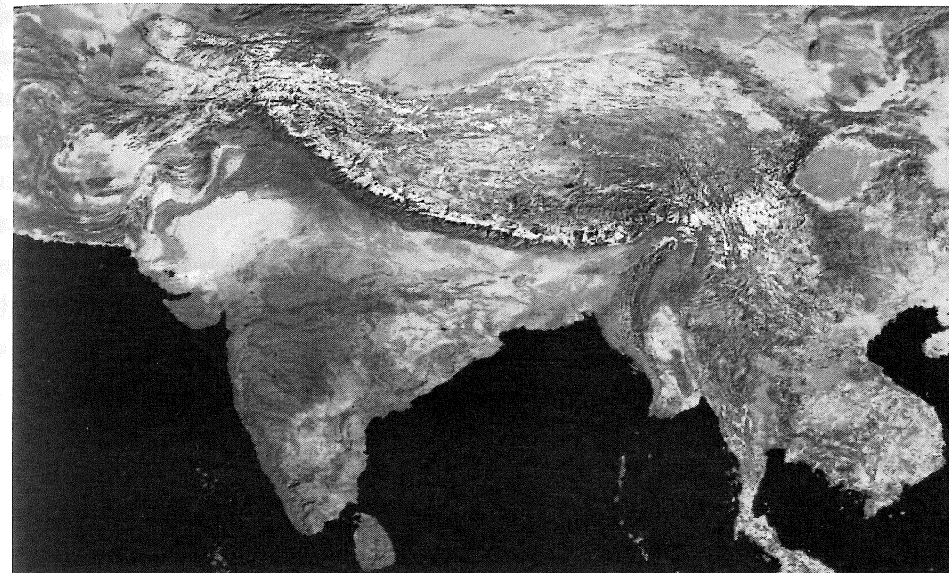
Director of the Himalayan Languages Project

Linguistic Population Prehistory of the Greater Himalayan Region: Interpretations of Emergent Genetic Data

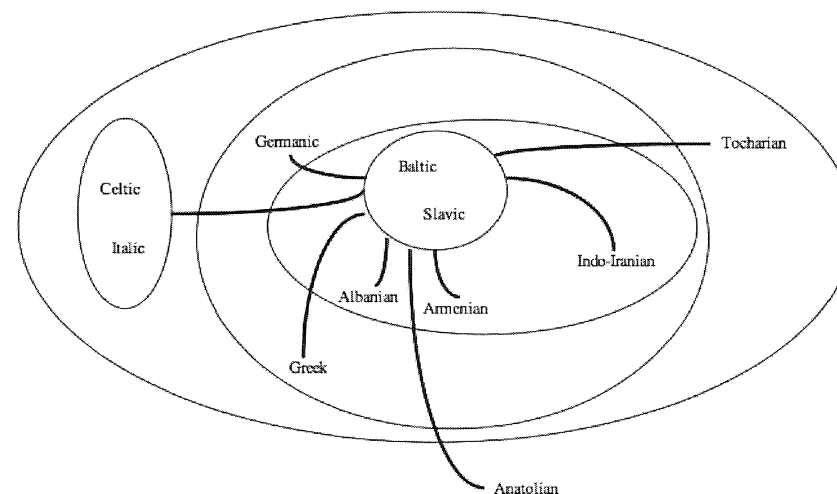
I will talk about the ethnolinguistic population prehistory of the greater Himalayan region, where I have been working for some twenty-five years as a linguist and, in recent years, together with population geneticists on the findings of our joint population genetic studies. So, what do these genetic studies tell us? As a linguist, I want to talk about that.



Now, behold the area. The Himalayas are the largest land barrier on the face of the planet, and the highly complex topography is also paralleled by a very complex ethno-linguistic mosaic. Around the Himalayas you find six major language families or linguistic phyla: (1) Austroasiatic, which used to be known as ‘Mon-Khmer-Kolarian’, (2) Dravidian, (3) Kra-Dai, which is also found in the literature under the older names ‘Tai-Kadai’ or ‘Daic’, and then (4) Altaic, the Altaic language family, (5) Tibeto-Burman – you will also find the older names ‘Indo-Chinese’ or ‘Sino-Tibetan’ for the language family, denoting obsolete phylogenetic models –, and finally (6) Indo-European. So, you have six large language families, and then you have a number of isolates (I don’t have a pointer): Kusunda is spoken in this area, and Nahali is spoken in this area. Burushaski is spoken in this area, and Vedda is still spoken in some parts of Sri Lanka. Andamanese is here, on these islands, and actually perhaps, Andamanese represents two distinct language families, and so there is considerable heterogeneity, of course, even within this area.

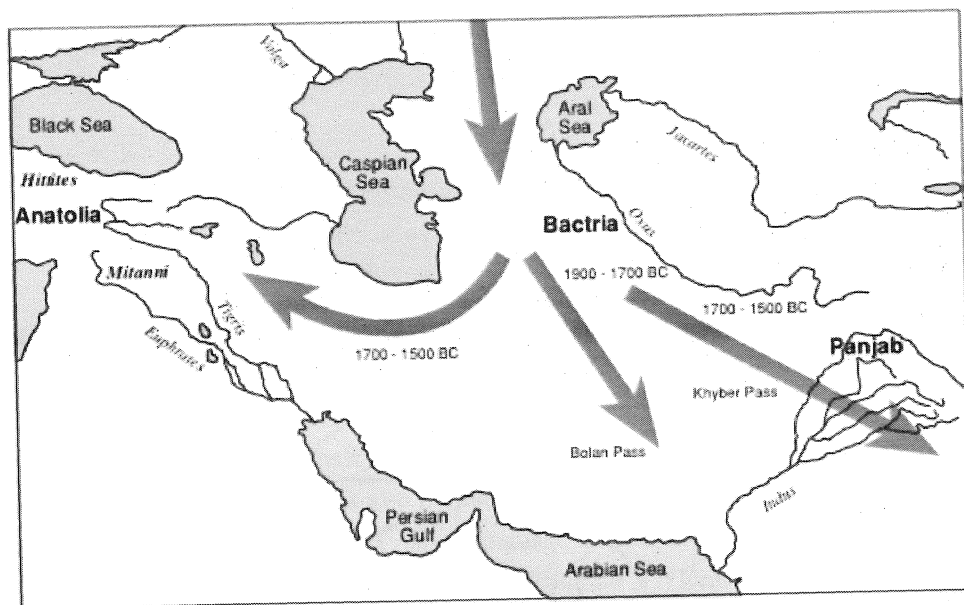


So, we have a number of minor language families, a number of isolates and six major language families. This mosaic yields a highly complicated ethno linguistic picture, and the Himalayas themselves are the very place where the two most populous language families on the planet meet: Tibeto-Burman and Indo-European meet exactly here. And so what is happening there, and what we can say about it as linguists working together with population geneticists?



Now, if we look at the Indo-European, then this is a ‘fallen leaves’ model of the Indo-European language family. Here we look at the various branches, and we do not impose upon the tree any higher-order subgrouping. When we look at these major branches of the family, what

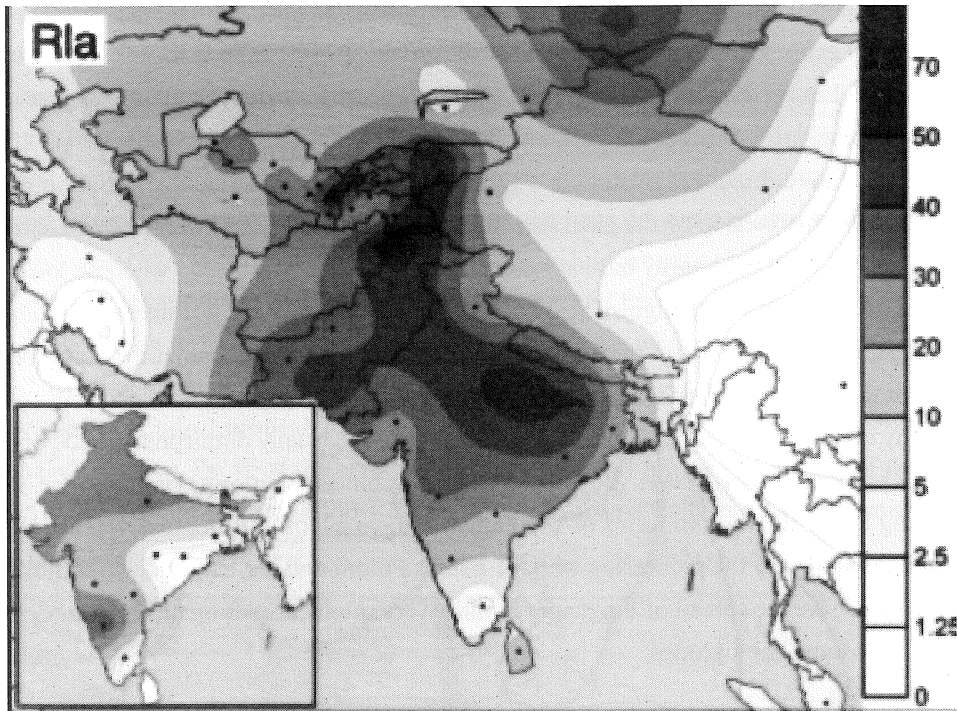
you notice is that half of the branches of Indo-European that we know about are extinct. And of those branches that are not extinct, half of those languages in those subgroups, those surviving branches, are also extinct. For example, Italic comprises Faliscan, Oscan, Umbrian, Latin and some languages which we might or might not know about, and only one of these languages, Latin, is alive today in the form of its modern dialects French, Romanian, Portuguese and so forth. Likewise, in the other branches you have the same situation where most of the languages that we know about in this language phylum are extinct. Ultimately, we have very few languages surviving of all the known languages and branches of the family. So why should the situation be different in other language families for which we do not have written records? We may imagine that many languages everywhere have gone extinct. Only one of those branches of Indo-European that we looked at in the previous diagram is represented in the Himalayan region. So, all of the branches of Indo-European are located outside of the Himalayan area. Only one of them, Indo-Iranian, is represented partially in the Himalayan area.



The scenario illustrated here is widely presumed. In fact, linguists presumed this course of events long before archaeologist ever found the Harappan civilisation or the Bactria Margiana archaeological complex or any related cultural assemblages. That is, before we ever knew about an Andronovo Bronze Age culture, linguists had already proposed this type of ancient population movement on the basis of textual and linguistic evidence. Some people falsely suppose nowadays that the archaeology antedates the linguistics. However, in fact linguists and archaeologists have been relating the original ethnolinguistic model and the older linguistic hypotheses to the findings of subsequent archaeological research.

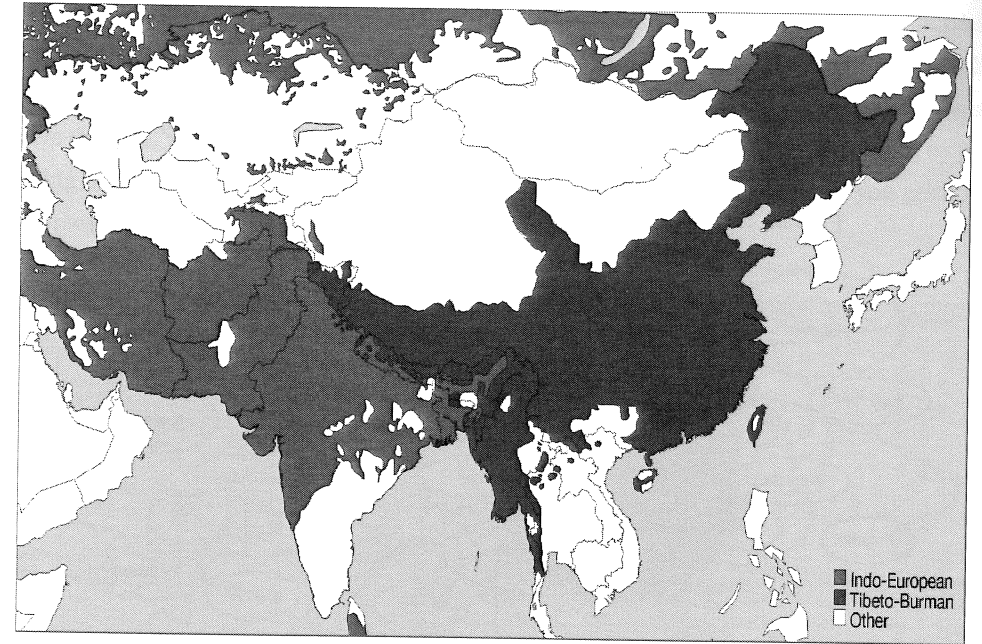
This brings up a question which I think was implicit in the previous speaker's talk but not made very explicit, although I think that he has made it quite explicit in writing in the past, and so do I in my handbook *Languages of the Himalayas*. When we are looking at the archaeology, we are looking at the past, but we are looking at just one version of the past, which is the material culture. Linguistics also gives a version of the past, and population genetics gives us another version of the past, and these three versions of prehistory can be correlated but they need not necessarily have anything to do with each other. And when we look at the findings of genetics and linguistics, obviously genetic answers are not necessarily linguistic answers. What I have termed the Father Tongue hypothesis is implicit anytime that geneticists look at the distribution of Y chromosome haplogroups and languages. In such cases, you are already looking for a correlation and you are looking at sexually dimorphic genetic prehistories of populations. Whenever you look at Y chromosome haplogroups and language, you are looking at the situation in terms of the Father Tongue hypothesis because you are looking, based originally on the pioneering work of Estella Poloni and her team-mates, at those areas where you have the spread of languages families very often correlating nicely in tandem with Y chromosome haplogroups.

However, sometimes you see the very opposite happening. You see, for example, in Baltistan in what today is northern Pakistan the mitochondrial DNA being of the same type as the mitochondrial DNA that you find amongst all Tibetan speakers. The Balti people speak the most conservative dialects of Tibetan, or the most phonologically conservative Tibetan languages, but their Y haplogroups are from the Near East, as is their religion. So there may be a correlation between paternal lineage and religion, but the linguistic correlation is not with the paternal lineage but with the maternal lineage. The prehistory of any population may be sexually dimorphic. And then comes another third dimension of the archaeologically attested material culture. Not only do we have three versions of the past, but we have the problem of time depth. The time depth that you are looking at in these three versions of prehistory may be very different indeed. So if you look at the Y-SNP data, i.e. the single nucleotide polymorphisms on the Y chromosome, for India then we see haplogroups identified there correlating possibly with ancient Indo-Europeans entering the Indian Subcontinent from the northwest. The mitochondrial DNA lineage in India based on the work of Toomas Kivisild and his co-workers as well as the findings of other research groups look indigenous. So we actually get the picture of men coming in with their weaponry, as recounted in the Rgveda, where we have male bands marauding and attacking as *puramdara*, destroyers of the native walled Indus cities. So the genetic picture actually correlates satisfactorily with the picture painted by philologists and linguists on the basis of textual sources and linguistic evidence.



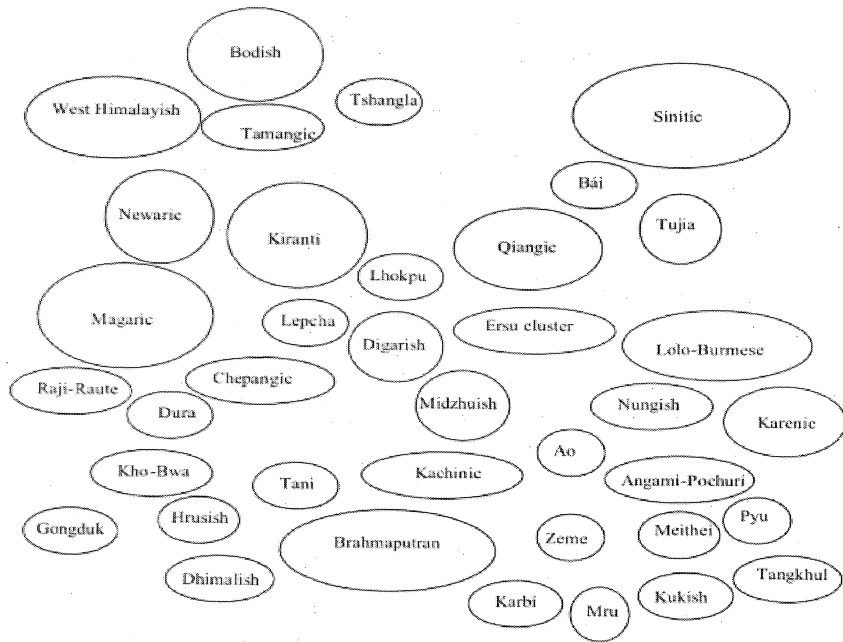
*Spatial frequency distribution map of haplogroup R1a (Sahoo et al. 2006).
Caste populations and national majority groups shown on main map, frequency gradients
amongst tribal populations on the inset map.
(You can see this figure in colour further)*

And now because our kind hosts are from Kazakhstan, you see that this gradient comes in, and it is a very strong gradient. The frequency gradient of the relevant Y haplogroups appear very well to represent Indo-European seemingly coming in from the northwest, but the Y haplogroups in question are not very strong in Central Asia, and of course there is a reason for that. That is because of subsequent genography, but this Y chromosome gradient is presumed or argued to be connected with and to be coetaneous with the advent of Vedic Sanskrit speakers into the Subcontinent.

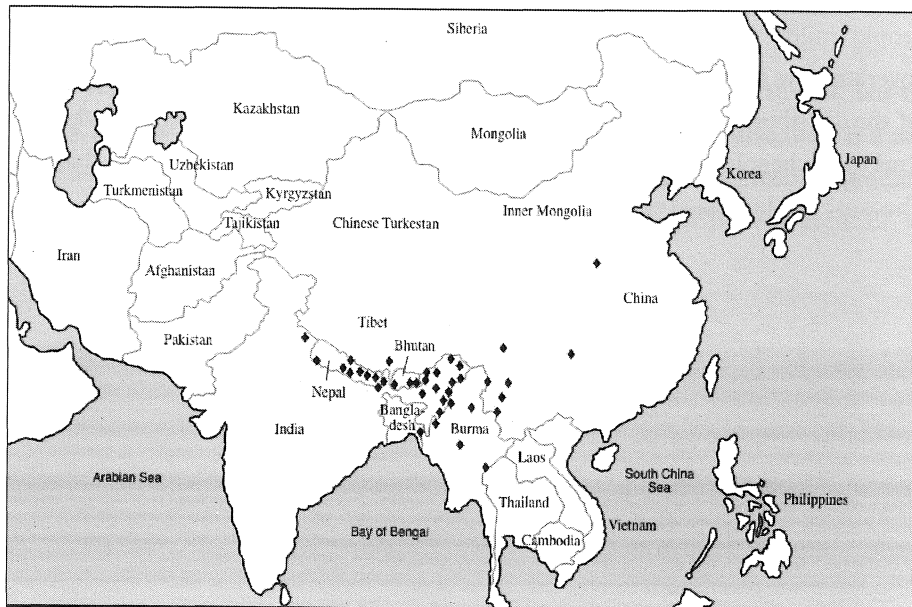


(You can see this figure in colour further)

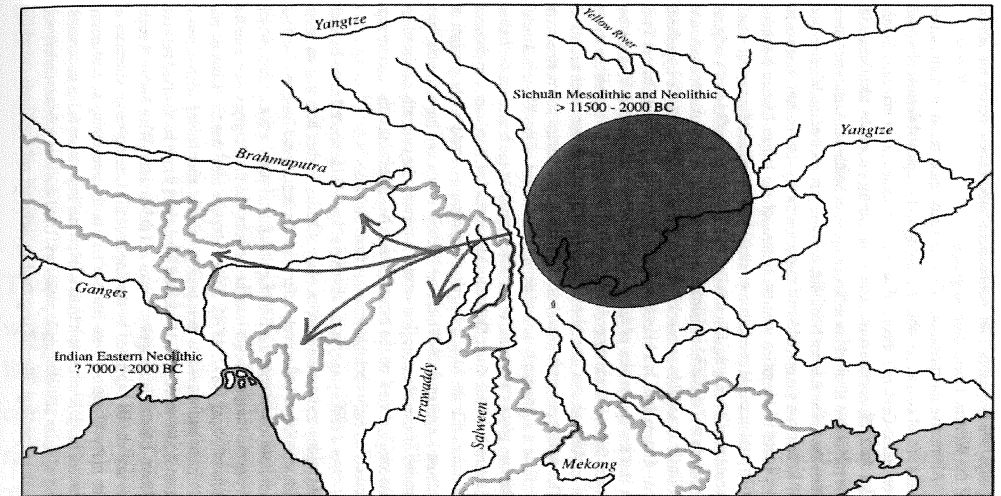
Now, if you look at the distribution of these, the two largest, not really the largest, but the most populous language families on the planet, then we see that this is where they meet. We see that these Tibeto-Burman are surrounded by the Indo-Europeans actually all of this area to the north in Siberia on the top of the map are Russian speakers. Historically, we know that this is the result of Ermak Timofeevič moving into Siberia and in the aftermath his people trailing in behind him. So, in addition to this historically recent colonial expansion overland, we see that the Chinese presence in Manchuria is also historically recent. The rest of the distribution of Sinitic is far older and of greater interest, especially when we look in terms of archaeology. When we look at Central Asia, there is a blank on this language map because the language communities belong to another language phylum.



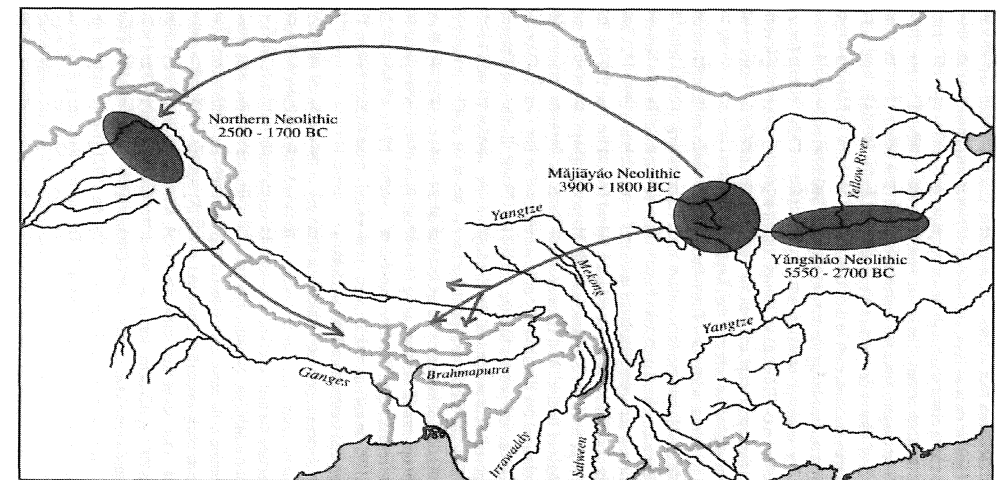
If we look at the major Tibeto-Burman groups on this diagram, again these major subgroups of 'fallen leaves' are not languages but subgroups or language branches. So each individual subgroup may contain forty languages or just one language or perhaps eight languages, but the 'fallen leaves' in the diagram represent major subgroups that are recognised by linguists to be very distinct one from the other. These are the major recognised subgroups. This slide shows their geographical distribution. So this is where you find these subgroups.



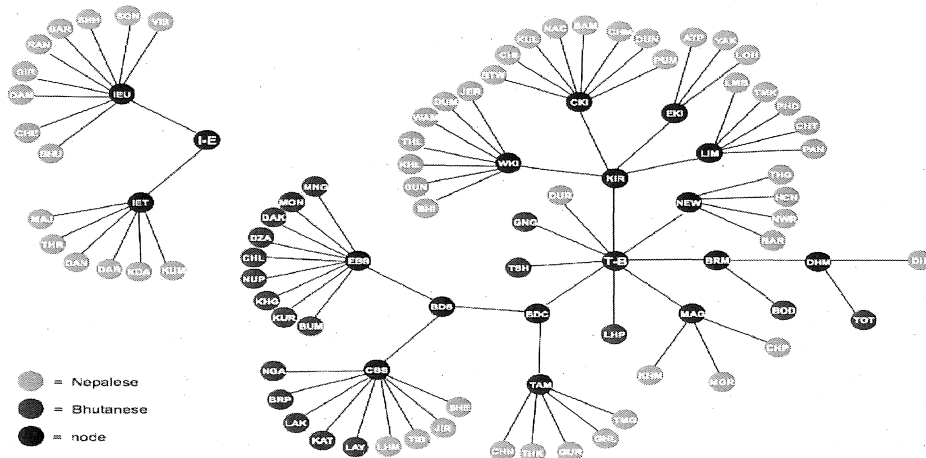
So if you think that languages may have died and gone extinct, in which area might they have died? When you are trying to reconstruct the prehistory of language communities, then what is the centre of gravity of the language family from the linguistic point of view? On the map, we see that this is distribution of major subgroups. Now, when you take this picture as your starting point and look at the findings of population genetics and archaeology, you get mixed information.



The spread of the Sichuan Neolithic into northeastern India is not really disputed, but the dates range widely because we do not have a very good stratigraphy or good dendrologically calibrated radiocarbon dates for northeastern India or any accelerator mass spectrometry dates for this portion of India, although now these dates are coming.

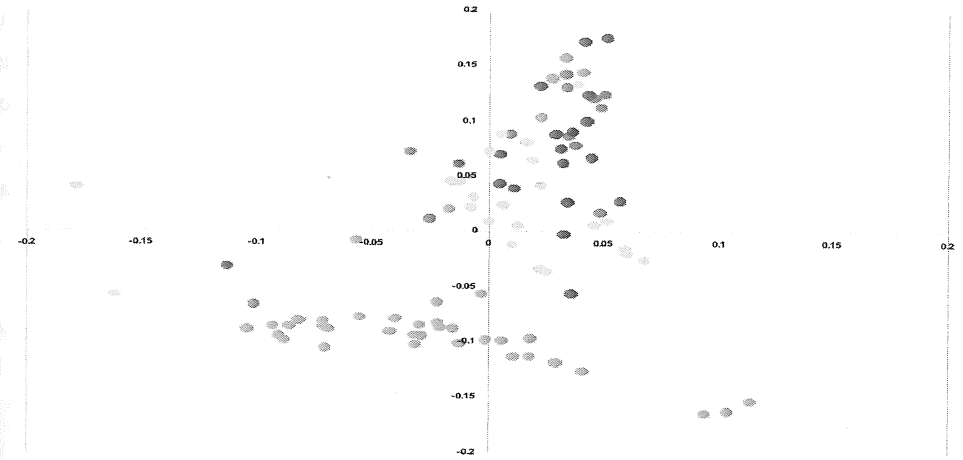


What you see here on the map is also a movement that is well attested in the middle of the third millennium BC. You have the Mǎjiāyáo culture having colonial exponents in western Tibet at Chab-mdo and at mKhar-ro and in Sikkim and also in Kashmir and moving into the Himalayan area. Although this Neolithic expansion gives us relatively recent dates, it does not account for the type of diversity we see inside the Himalayan region and the time depth that we are inclined to impute to this diversity as linguists as opposed to the relative lack of diversity that we see outside of the Himalayan region to the north and to the east.



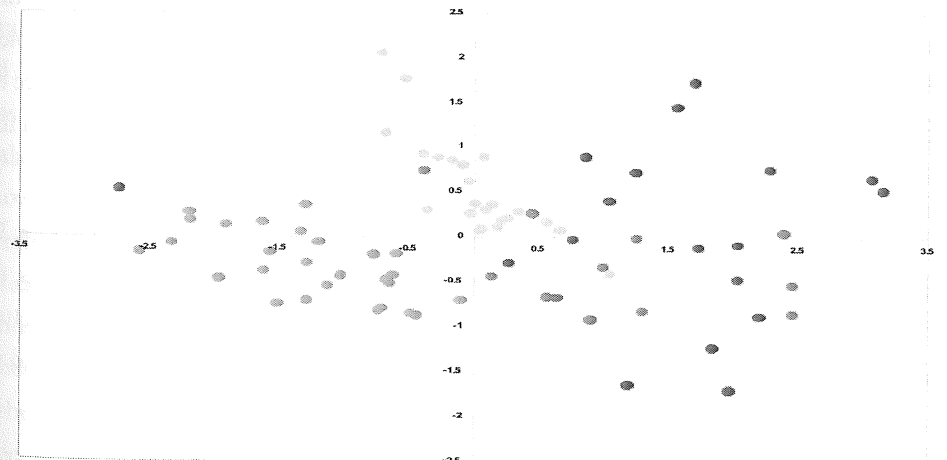
(You can see this figure in colour further)

Now we have taken genetic samples from all of the major linguistic communities in Nepal and in Bhutan, and we have compared them also with the data that we have from India and China. We have put these findings into the network diagram shown here. This is a linguistic tree or Stammbaum adapted in the shape of a network for the consumption of geneticists so that we can speak with each other in a meaningful way. You see Tibeto-Burman and Indo-European on separate networks because they are separate language families. Then this is what you get if you plot the Y-SNPs, reflecting the parental lineage, in a multi-dimensional scaling diagram.



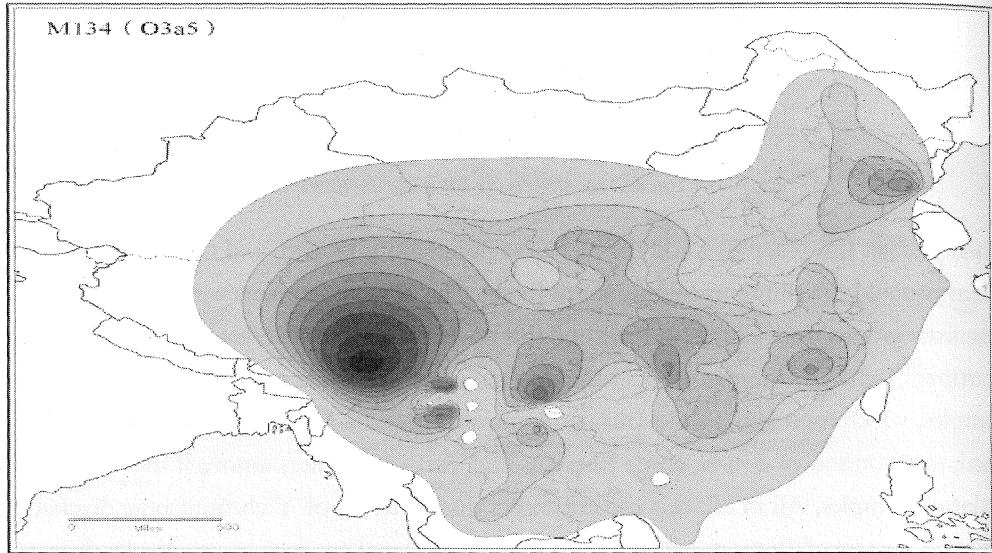
(You can see this figure in colour further)

Then you get this when you add the Nepalese populations [green], and this is what happens when you add the Bhutanese populations [blue]. This is what happens when you add ethnolinguistic groups from China [yellow], and this is what happens when you add the Indian subcontinent [pink] to the picture. On the diagram, you see some outliers within the Indian samples, which represent Tibeto-Burman groups from Indian subcontinent such as the Apa Tani, and you see, of course, some Russians and Turks as outliers amongst the Chinese population samples. All in all, this picture shows that in terms of Y chromosome diversity, a country the size of Bhutan and a country the size of Nepal are each as genetically diverse as the whole of China or the whole of India. What we also know from looking at mismatch distributions is that this genetic diversity in the Himalayan heartland is quite old.



(You can see this figure in colour further)

In the next diagram, we are looking at the mitochondrial DNA, and in particular at the hypervariable sequence or hypervariable region. Now, this is the picture when we add the Bhutanese populations to the Nepalese populations sampled, then when adding China and now when adding India. The diagram shows that the Himalayan area is highly diverse and, moreover, that the Himalayas as a genetic region do not really occupy a position intermediate between the two.

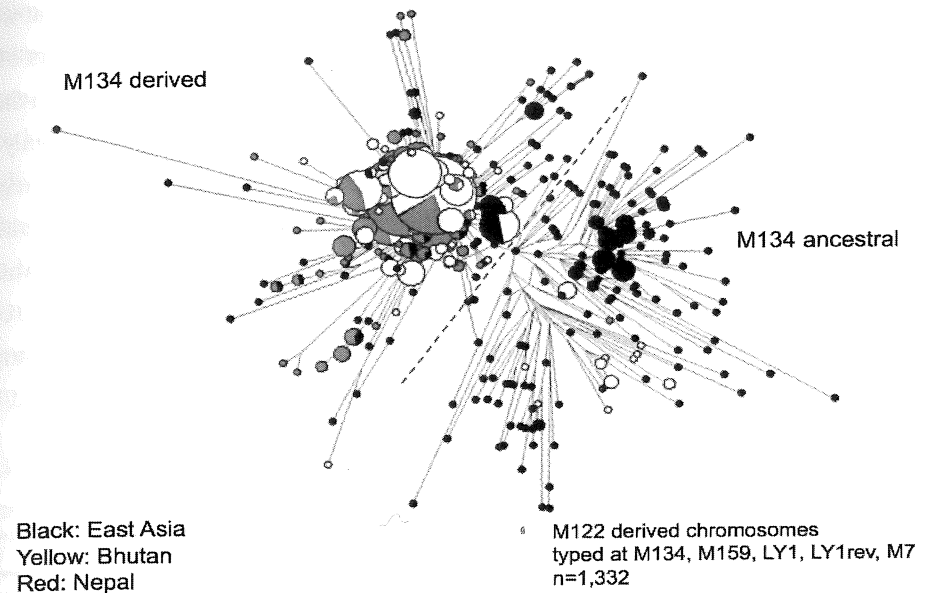


Spatial frequency distribution of Y haplogroup O3e (M134) (Shi et al. 2005)

Next we look again at the Y chromosome haplogroup that we see as being best represented amongst speakers of Tibeto-Burman languages. The diagram shown here is from an older publication, which does not yet contain our data from the Himalayan region. Here we see the epicentre of the frequency gradient of this Y chromosome haplogroup based on the data that were available to these researchers.

Meanwhile, our research programme in the Himalayan region has accrued a lot more data. We have a lot more data from areas which were never sampled before. Based on our findings, the epicentre of this frequency gradient moves considerably from where it is now on this picture. The next picture shows a rooted topology for the haplogroup in question, but where do we draw the line?

13 usat (DYS19, DYS388, DYS389I, DYS390, DYS391, DYS392, DYS393, DYS426, DYS434, DYS435, DYS436, DYS437, DYS438)



(You can see this figure in colour further)

That line does not really correspond to the Himalayas. What we have found out is that the geographical distribution of the major Tibeto-Burman subgroups indicates – if we may for the sake of argument assume some relationship here between the two – that there is a clear-cut correlation between the genes and the languages. Yet, of course, as we have so often stressed, the linguistic ancestors and biological ancestors of any given language community need not have been the same people in prehistory, as we have seen demonstrated in many areas, the Hungarians being a case in point, where linguistic and biological ancestry is clearly not well correlated. Yet, in the case of Tibeto-Burman, the homeland of the language family must have lain – as long as we assume here for the sake of argument the correlation within the greater Himalayan region based on the analysis of the polymorphisms on the Y chromosome, whilst knowing the frequency gradient of for the relevant Y-SNPs – the homeland of the language family as a whole, as this paternal lineage suggests, must have lain to the north. What the genetics of modern Tibeto-Burman language communities shows is the persistence of thousands of years of stable isolated habitation in the Himalayan region. So ancient Tibeto-Burman language communities lived in the Himalayan region for a long time undisturbed, maybe not always peacefully but co-existing there relatively undisturbed for a long time. When we date this, inasmuch as geneticists are able to date this using coalescence times by what is believed by some to be a molecular clock, or what at least has been called a mole-

cular clock, then the period of ancient Tibeto-Burman habitation is datable roughly to the time of the last glacial maximum, which makes it very old. This is decidedly a time depth at which linguists traditionally feel rather uncomfortable. Another problem is the precise extent of the glaciation during the last glacial maximum in the Himalayan region because, although the extent of the ice is not disputed in areas such as Europe, the extent of glaciation is highly disputed and debated by palaeoclimatologists and palaeontologists in the Himalayan region. The hitherto recoverable findings have been interpreted both ways in the specialist literature, with palaeoclimatologists divided into two camps. Whether the Himalayan region was actually a refuge area containing many pockets and shelters within the Himalayas with the complex myriad of local microclimates, or whether the Himalayan heartland was largely uninhabitable during the last glacial maximum... This issue is something that is disputed, and there is a large body of palaeoclimatological literature disputing this very issue.



In keeping with the theme of this conference, I am deviating now to Central Asia. But this is an altogether different story. The genography of Central Asia is perhaps a bit different. At least my friend Peter Underhill, recently at an Ethnogenesis Roundtable held at Harvard University, maintained that the genography of Central Asia was like that of Manhattan. That is to say that we have very recent genography in Central Asia. I shall not argue here whether or not his contention is true. Rather, if that were to be true, historians and linguists would not be terribly surprised because we know that within the Altaic language family, Turkic is speci-

fically one branch of Altaic which at some point spread into Central Asia relatively recently. Presuming that the Indo-Europeans preceded them, of course, then the early Turkic language communities in this area would have had to come after that. Here are some dates for you. Therefore, we might not expect, if this historical scenario is correct and if Turkic indeed represents a recent incursion into Central Asia, then we might not expect the geographical distances to correlate necessarily as well as with the genetic and linguistic distances as they appear to do in the Himalayas, where genetic distances do correlate extremely well with linguistic distances between the major language families, though not necessarily for major subgroups within these families at our present state of knowledge. Yet the population genetic story of Central Asia is now beginning to unfold, and time will tell whether or not many earlier contentions are borne out by the results.

In closing, I would like to make some final comments pertaining to my personal observations of somatology. Once I used worked in the Soviet Union and also subsequently, that is ever since I naturally continue to meet people from this vast area. What might you expect if people were to have entered this area so quickly from the Altaic region is that they could have borne older genetic splits into the region, and the date of the genetic differentiation would then be much older than the dates of their arrival into the geographical area, and that these older genetic splits could have been introduced by them as they settled. So, I think that we shall see some similarities but mainly also some contrasts between the Himalayan region and Central Asia. Yet we cannot yet know what we shall be able to say with certainty because, given the distribution of Turkic languages in Central Asia, we shall still have to consider three different types of evidence, coming in from population genetics, archaeology and linguistics. We get three different pictures of the past, and we get these different versions of prehistory at three different time depths. How well can we correlate these pictures, and what kind of ancient Tibeto-Burman languages may meanwhile have vanished in what today is China, especially with the spread of Chinese or different varieties of Sinitic emanating from the ever shifting imperial centres? Many ethnolinguistic groups are listed in the Chinese chronicles of which there is no linguistic trace left. And so, what was the direction of assimilation, and what was the direction of movement? Looking from the Himalayas, we see all the diversity nearby, and Chinese or Sinitic we see to be an outlier.

So, we see the highest language diversity within the Himalayan region, with languages and linguistic subgroups differing from each other most extremely within the Himalayan area itself, in fact with particular subgroups such as Gongduk often differing more one from the other, then that certain individual Tibeto-Burman subgroup differ from Chinese. So for Tibeto-Burman we have the greatest diversity linguistically within the Himalayas, and also

great genetic diversity can be observed there too. So I am not going to give you the answer to the conundrum here, but the solution must be one which recognises the different layers in population prehistory. Obviously, in interpreting the prehistory of the region, we shall have to contend with the emergent and increasingly fine-mesh genetic data. Until the occasion presents itself for us to unveil the entire picture unfolding in all of its complexity, I am going to let you mull over the implications of the three different pictures of the past presented here in light of the contested higher-order subgrouping hierarchy and the likely past and present geographical distribution of the Tibeto-Burman linguistic mosaic, the chronology and patterns of dispersal of identifiable cultural assemblages in the archaeological record, and the complex and sexually dimorphic population genetic prehistories which are now unfolding. That's all.

Vera Budanova

*Professor at the Institute of the World History and Archives,
Russian State Humanitarian University (Moscow)*

The Great Migration: Ethnic Perspective of the Era of Change

The entire history of humanity confirms that man on Earth finds himself in a permanent state of constant migrations. However, historians make a distinction between the «great migrations of peoples» as the stages of the highest increase in the migratory activity in any given region in the world, and «The Great Migration» in Europe and Asia as a specific period between the antiquity and the middle ages, when for the first time the migratory process becomes part of a complex, varied and dynamic interaction between Barbarism and Civilization, and reaches a most intensive phase of the opposition of the center of civilization to its barbarian periphery. The paradox, however, is in the fact that although migrations of the II - VI cc. proved to be an external and alien element for civilization, they were to a considerable extent generated by civilization itself. Both the tribal associations and civilizations do not in themselves explain the reasons for migrations. In my view, not excluding the climatic, demographic, economic and other reasons, closest attention should be paid to the process of interactivity between civilization and the tribal world beyond its boundaries. Migratory pulsations and the growth of civilizations are interdependent. Migratory flows stretch out not only towards the large territorial expanses, but first and foremost towards the vital expanse offered by the centers of civilization. Feeling the ruinous influence of an «overdeveloped» and decomposing society, the tribal world enters the migratory process, becoming the instrument of death for civilization itself.

Traditionally The Great Migration was considered uniquely as a European phenomenon, as a symbol to define the passage from the antiquity to the middle ages. However, overcoming Eurocentrism, it must be noted that the European «model» (Barbarikum – the Roman empire), of The Great Migration constantly changes, influenced by and periodically intersecting with the Asian «model» (nomadic tribal world – the Han Empire). For this very reason we may justifiably consider The Great Migration as part of world history, as a unique transitional period. It is during a significant period in history (no longer the antiquity, but not yet the Middle Ages), and limited to a precise chronological period (II -VII cc.) and to a specific territory (Europe, Asia, Africa), that the interaction of barbarism and civilization achieves its highest level and results in the creation of a new civilization – Medieval.

It is well known, that the attitude of the Roman and Chinese civilizations towards the barbarian world was from the outset that of opposition «We – They.» The very concept «Barbarians» played here a crucial role. Thus, the semantics of the Ancient World was revealed within