Unfortunately, they have not received much attention in linguistics, even though the idea of quantitative trees was suggested by the linguist August Schleicher over a century ago. Later in this paper, we shall attempt to quantify some linguistic trees.

In cases where the time depth is not great, and where there is adequate supporting textual material, the genetic relations are not difficult to establish. This is the case of the Chinese dialects, which began their differentiation some 2000 years ago. Similarly, the Germanic languages seem to have a comparable time depth. In Chinese, the written materials actually pre-date the differentiation into the modern dialects by a considerable margin. For Germanic, the earliest materials go back some 1700 years. Such materials provide an important source of data for studying genetic relations.

Even for languages with long literary traditions, written materials do not ever cover time depths to the degree that linguists wish to investigate. The further back in time we go, the scantier the data become, and correspondingly, the less robust the linguistic hypotheses. It becomes increasingly imperative that linguists should coordinate their work with other disciplines that can help shed light on the distant past of human civilization.

Sharing a past could result in resemblances of many different types and forms. It could result in similarities in painting and sculpture, such as those noted for early peoples on both sides of the Pacific. It could result in similarities in belief systems, such as those involving the tiger among certain Miao tribes of Sichuan.

As a third example of cultural resemblance, we may consider the "fuzi lamming shi" or patronymic linkage system. This is a system for showing generational order and affiliation, "whereby the name of a son always contains an element from the name of his father, e.g., Sheng-lo-p’i, P’i-lo-ko, Ko-lo-feng, Feng-ch’i-hi, and so on."

As a key cultural feature, according to Backus, which determines that the rulers of the ancient Nanzhao kingdom in Tang-Song China were a Tibeto-Burman people, rather than of Thai lineage, as had been claimed for several previous decades.

The three windows in the title of this paper refer to three
disciplines that have been systematically brought together for investigating prehistory, in the works of some leading scholars studying these questions.8 These are anthropology, in particular archeology and physical anthropology, genetics, in particular genetics of human populations, and evolutionary linguistics.9 In the remainder of this paper, I will discuss briefly the light that each of these windows has shed on China’s distant past.

1. Archeology

In terms of archeology, the most recent “news” which has caught the attention of the media in the West has to do with a large collection of desiccated corpses found in eastern Xinjiang.10 These finds are due to the dedicated search of Wang Binghua and other archeologists, starting in the late 1970s. Some of these corpses date as far back as 4000 years, though others are considerably younger. Because they are remarkably well preserved by the desert climate of Xinjiang, they are clearly recognizable as Caucasian.

While the magazine cover may lure readers with the question “What are 4000-year-old Caucasians doing in Xinjiang?”, scholars have long known about the important routes connecting the Central Plains [中 ] of China with the rest of Eurasia through the Gansu corridor and Xinjiang.11 Intensive interactions between the peoples of the

---


9The more usual term, “historical linguistics”, is too limiting as the field is currently practiced by many investigators. In addition to the detailed reconstruction of well-accepted family relationships, the field should be concerned as well with larger issues of even greater importance and interest, such as deeper relations among these families, and developing methods for quantifying and dating these relations. In these later respects, evolutionary linguistics will find much that is useful in the advances recently made in evolutionary biology. A recent effort to address such larger issues is that of Ruhlen, Merritt. 1994. *On the Origin of Languages: Studies in Linguistic Taxonomy*. Stanford University Press.

10Hadingham, Evan. 1994. The mummies of Xinjiang. *Discover* 15:68-77. This popular article contains a series of striking photographs. These archeological finds are being effectively publicized by Victor Mair of the University of Pennsylvania. Since the writing of these pages, Mair has provided additional discussion of these finds in his *Mummies of the Tarim Basin* *Archaeology* 48:28-35, 1995. See also the *Journal of Indo-European Studies* 23:3/4, Fall/Winter 1995 for an interdisciplinary collection of writings on this topic.

11See, for instance, the detailed observations in Pulleyblank, E. G. 1966.

---

Central Plains with those further west have been copiously recorded in Chinese history books, so despite its commercial allure the question is not really all that new or startling.12

On the other hand, we are deeply interested in the language(s) these early inhabitants spoke, even though clear answers may not be forthcoming for some time. An immediate association is with Tocharian,13 a language that has long been extinct. These materials, written in an Indian Brahmic script during the Tang dynasty, were discovered in Xinjiang around the turn of the century. The language is now universally accepted to be Indo-European. According to Bardi, “it is now generally held that the speakers of Tocharian were part of a very early migration from the Central Indo-European area, possibly as early as 2000 BCE.” 14

If we accept this statement, then we have a match between the archeology and the linguistics not only in place but also in time. But it is not clear what evidence lies behind the date of 2000 BCE which Bardi accepts; perhaps he is following the reasoning that Pulleyblank offered; see footnote 11. Furthermore, it is known that Iranian languages had a much greater sphere of influence in Xinjiang.15

---


9The more usual term, “historical linguistics”, is too limiting as the field is currently practiced by many investigators. In addition to the detailed reconstruction of well-accepted family relationships, the field should be concerned as well with larger issues of even greater importance and interest, such as deeper relations among these families, and developing methods for quantifying and dating these relations. In these later respects, evolutionary linguistics will find much that is useful in the advances recently made in evolutionary biology. A recent effort to address such larger issues is that of Ruhlen, Merritt. 1994. *On the Origin of Languages: Studies in Linguistic Taxonomy*. Stanford University Press.

10Hadingham, Evan. 1994. The mummies of Xinjiang. *Discover* 15:68-77. This popular article contains a series of striking photographs. These archeological finds are being effectively publicized by Victor Mair of the University of Pennsylvania. Since the writing of these pages, Mair has provided additional discussion of these finds in his *Mummies of the Tarim Basin* *Archaeology* 48:28-35, 1995. See also the *Journal of Indo-European Studies* 23:3/4, Fall/Winter 1995 for an interdisciplinary collection of writings on this topic.

11See, for instance, the detailed observations in Pulleyblank, E. G. 1966.

---

Central Plains with those further west have been copiously recorded in Chinese history books, so despite its commercial allure the question is not really all that new or startling.12

On the other hand, we are deeply interested in the language(s) these early inhabitants spoke, even though clear answers may not be forthcoming for some time. An immediate association is with Tocharian,13 a language that has long been extinct. These materials, written in an Indian Brahmic script during the Tang dynasty, were discovered in Xinjiang around the turn of the century. The language is now universally accepted to be Indo-European. According to Bardi, “it is now generally held that the speakers of Tocharian were part of a very early migration from the Central Indo-European area, possibly as early as 2000 BCE.” 14

If we accept this statement, then we have a match between the archeology and the linguistics not only in place but also in time. But it is not clear what evidence lies behind the date of 2000 BCE which Bardi accepts; perhaps he is following the reasoning that Pulleyblank offered; see footnote 11. Furthermore, it is known that Iranian languages had a much greater sphere of influence in Xinjiang.15

---

Chinese and Indo-Europeans. *J. of Royal Asiatic Society* 9:39. “The nearest Indo-Europeans to China in historical times were the Tocharian-speaking inhabitants of the oases on the northern rim of the Tarim basin. The linguistic position of Tocharian makes it probable that its speakers always lay to the east of the Indo-Iranians. This in turn implies that they probably arrived on the western borders of China not later than the time of the Aryan invasion of India in the second millennium BCE. If we posit an Asian origin for the Indo-Europeans, their arrival was probably earlier.” P. 35.


13The name Tocharian itself has not gained consensus among scholars. See the remarks by G. S. Lane on the present state of Tocharian research in the *Proceedings of the VIII International Congress of Linguists* 252-266.


before they were supplanted by Turkic and Mongolian languages many centuries ago. There is no reason to believe that all the unearthed individuals spoke the same language; if anything, the contrary is more likely. At any rate, genetic studies on these corpses will eventually shed more light on their affiliations.

Human activity on the land that is now China goes a great deal further back than a mere 4000 years. According to the authoritative survey of K. C. Chang, stone tools date back some 2 million years\(^{16}\) [p.35], and the earliest pottery dates to approximately 9000 BP [p.105]. This latter date, which marks some rudimentary form of agriculture, has more than twice the time depth of the Xinjiang Caucasians discussed above. It is also of interest to note that the site of this earliest pottery, Zengpiyan [甄皮岩], is in Guangxi in South China, many thousand miles away from the Xinjiang discoveries.

One of the most important fossil discoveries is the cranium unearthed at Dali [大荔] in Shaanxi, dated to 200,000 BP. This remarkably well-preserved cranium of a male less than 30 years of age is reported to have a cranial capacity of 1120 cc.\(^{17}\) This indicates a brain size well within the range of modern peoples; indirectly, such a brain size is highly suggestive of a capacity for language.

The number of human fossils increases significantly as we get closer to the present. An experiment was carried out on a series of head measurements made on fossils dated to the range of 4000-7000 BP.\(^{18}\) This experiment is pertinent for the light it may shed on the distribution of ancient peoples in China. These measurements were analyzed by average linkage, and the resulting tree diagram is shown in Figure 1, taken from Wu and Olsen, p.121.

\(^{16}\)Chang, K. C. 1986. The Archaeology of Ancient China. 4th ed. Yale UP. Some of the earlier dates are not generally accepted. Referring to a paper by Y. C. Tang et al in *Vertebrata Palaeo-assatica*, 1981, on which the date of 2 million is based, Desmond Clark writes me: "The paper uses the paleo-mag. column obtained for the classic geological sections at Haojiatai and suggests where, in the sequence, the Xiaochangliang site fits. Since Haojiatai is more than 12 km from Xiaochangliang and there are a number of small faults between the two localities, I would not attach much credence to a correlation of this kind." [Personal communication, April 14, 1994].


\(^{18}\)The experiment is reported by Wu Xinshi and Zhang Zhenbiao in Wu and Olsen [1985:107-193]. The statistical analysis used in the experiment, average linkage, is subject to certain limitations which will be discussed later in the paper. In particular, the analysis assumes a constant rate of development along every lineage, which is not realistic. Further, it is important to buttress the data with additional data from a larger number of sites than the nine in Figure 1. Nonetheless, the figure clearly has a suggestive value.

---

*Victor H. Mair, editor*
east and west, with Yi and Shang belonging to the eastern system while Xia and Zhou belong to the western system. If we look at China's development on a grander time scale, however, we see that a more basic division is actually between the north and south. This division is seen as early as in Neolithic times, as in Figure 1 above, and persists to the present day, as revealed in genetic studies to be discussed below.

Actual language specimens, of course, are not available until somewhat later. It was a major achievement of archeology during the early decades of this century to uncover thousands of shells and bones, inscribed primarily for divination purposes over 3000 years ago. Enough knowledge has accumulated in their study so that there is no doubt that the texts etched onto these materials are in a language that is directly ancestral to the Sinitic languages.

The continuity in the shapes of the written characters as well as in a basic stock of words and phrases over these 3000 years is obvious. We are also beginning to inquire into similarities in syntax. An intriguing question, for instance, has to do with the evolution of interrogative structures of the A-not-A form found in many Chinese dialects. Such structures, quite distinctive in syntactic typology, are constructed by conjoining an affirmative sentence to its negative counterpart, with subsequent deletion of redundant materials. Thus, as shown below, [1] and [2] are conjoined to form [3], which in turn underlies [4] and [5].

[1] Ta yao chi mian /He wants to eat noodles
[2] Ta bu yao chi mian
[3] *Ta yao chi mian Ta bu yao chi mian
[5] Ta yao bu yao chi mian?

It is well known that the typical syntax of asking questions in divination was to inscribe an affirmative sentence on one side of the plastron and its negative counterpart on the other side. Sometimes the same plastron is used for a series of such A-not-A questions. Figure 2 shows five such pairs on a single plastron, as illustrated by Chou. Clearly, much intricate research needs to be done to demonstrate the


Victor H. Mair, editor

Three Windows on the Past

Three Windows on the Past

The Bronze Age and Early Iron Age Peoples of Eastern Central Asia

evolutionary sequence in detail. Nonetheless, the similarity across such a large time span is quite striking.

Figure 2: Oracle bone showing 5 pairs of affirmative-negative sentences. [From Chou, 11.11. 1979]

2. Genetics

Material remains from long ago, carefully interpreted in the perspective of archeological knowledge, can tell us much about the past. Yet there are other indicators of the past that do not lie buried under some silent mound awaiting discovery, but which are constantly within each of us today. This is our biological heritage, expressed in our genes. Charles Darwin was perhaps the first to connect this to
language when he wrote the following in chapter 14 of his *Origin of Species*:

If we possessed a perfect pedigree of mankind, a genealogical arrangement of the races of man would afford the best classification of the various languages now spoken throughout the world.

Kroeber was thinking in a similar direction when he remarked that

While populations can learn and unlearn languages, they tend to do so with reluctance and infinite slowness, especially while they remain in their inherited territories. Speech tends to be one of the most persistent populational characteristics; and "ethnic" boundaries are most often speech boundaries.25

Regarding persistence of language, or linguistic loyalty, one is reminded of the rime admonition of the Kejia people in China: "Sell your ancestor's land, if you must, but do not forget your ancestor's speech!"24

On the other hand, a page later in the same chapter, Kroeber reminds us that "Speech and culture have an existence and a continuity of their own, whose integrity does not depend on hereditary integrity. The two may move together or separately." In the Chinese setting, where the boundaries between ethnic groups have always been fluid and ever-changing, it turns out that the two, i.e., genes and language, often move separately.

Given that typically people do pass on both their genes and their language to their offspring, we should expect a strong correlation between the two phylogenetic systems. However, various factors significantly complicate the picture. Although the correlation should obtain in the default case, where the two move together, the interest is all the greater in cases where the two move separately, since they tell us that additional factors have entered the history which we need to sort out.

A special instance of such factors is when the marriage tradition is multiply exogamous, such as that reported for the Vaupes in Southeastern Colombia.25 When the parents do not share a native language by design, and when adults living in close proximity with each other speak half a dozen or more different languages, we can expect the genes and the language to go separate ways.

The situation is similar for the contexts in which pidgin languages arise. There also, the child of immigrant laborers has no clear model upon which to build its own language, being in an environment where several immigrant languages exist precariously alongside the colonial language. Such situations, as well as cases of multiple exogamy, pose interesting challenges to any historical linguistics that is limited to the tree diagram for portrayal of language relationships.

More generally than the above special situations, one can easily come up with numerous examples of Native Americans whose indigenous languages have been replaced by colonizing languages from Europe,26 where again the genetic lineage departs from the linguistic one. From China, there are cases which involve larger numbers. Consider the Manchu and the Hais, each with populations numbering in the millions. Yet each community has adopted dialects of Chinese, with little trace left of their original languages.

At present, the languages spoken in China belong to five major stocks: Sino-Tibetan, Austro-Asiatic, Altaic, Austronesian, and Indo-European. However, there is no consensus among scholars on the exact composition of these stocks. The numbers of their speakers may be seen in the following table of the ethnic groups.27 These groups, it should be noted, are established primarily in terms of the languages they speak, though the correlation between ethnic identity and language use is never perfect.28

---


24"Ning mai zuozn tian, bu wuzg zuozn yan" functionalism, not functionalism. To recognize the full force of this saying, one should recall the crucial value that land holds for an agricultural people.


26In her book, *Flutes of Fire* [Berkeley 1994], Leanne Hinton lists dozens of California Indian groups with no native speakers left and dozens more on the verge of extinction, pp. 27-33. It is not clear precisely how much genetic admixture has taken place with these groups, though one would expect a considerable extent.


28Some peoples in Guangxi identified ethnically as Yao, for example, speak a language which is closer to Miao; on the other hand, some peoples in Hainan identified ethnically as Miao speak a language which is closer to Yao. There are various ethnic groups in China who have essentially discontinued the use of their original language and switched over entirely to Han speech; an example is the Manchu, whose original language was a variety of Tungusic.

*The Bronze Age and Early Iron Age Peoples of Eastern Central Asia*
### Table: Population of ethnic groups, based on 1990 Census.

<table>
<thead>
<tr>
<th>LANGUAGE</th>
<th>SPEAKERS</th>
<th>OTHER NAMES</th>
<th>TOTAL SPEAKERS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. Hanitic</strong>&lt;sup&gt;29 30&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HAN</td>
<td>1,042,482,187</td>
<td>Huayu</td>
<td>1,069,906,345</td>
</tr>
<tr>
<td>MANCHU</td>
<td>9,821,180</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HUI</td>
<td>8,602,978</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>2. Tibetan</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZANG</td>
<td>4,593,330</td>
<td>Tibetan</td>
<td></td>
</tr>
<tr>
<td>ACHANG</td>
<td>27,708</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NU</td>
<td>27,123</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DULONG</td>
<td>5,816</td>
<td>Rawang</td>
<td></td>
</tr>
<tr>
<td>MONPA</td>
<td>7,475</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LIOBA</td>
<td>2,312</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>3. Yi</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>YI</td>
<td>6,572,173</td>
<td>Lolo</td>
<td></td>
</tr>
<tr>
<td>BAI</td>
<td>1,594,827</td>
<td>Minjia</td>
<td></td>
</tr>
<tr>
<td>TUJIA</td>
<td>5,704,223</td>
<td>Akha</td>
<td></td>
</tr>
<tr>
<td>HANI</td>
<td>1,253,952</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LISU</td>
<td>574,856</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LAHU</td>
<td>414,476</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NAXI</td>
<td>276,089</td>
<td>Moso</td>
<td></td>
</tr>
<tr>
<td>JINGPO</td>
<td>119,299</td>
<td>Kachin</td>
<td>16,526,746</td>
</tr>
<tr>
<td>JINUO</td>
<td>18,621</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>4. Qiang</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QIANG</td>
<td>198,252</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PUMI</td>
<td>29,657</td>
<td></td>
<td>227,909</td>
</tr>
<tr>
<td><strong>5. Miao-Yao</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MIAO</td>
<td>7,398,035</td>
<td>Hmong</td>
<td></td>
</tr>
<tr>
<td>YAO</td>
<td>2,134,013</td>
<td>Mien</td>
<td></td>
</tr>
<tr>
<td>SHE</td>
<td>630,378</td>
<td></td>
<td>10,162,426</td>
</tr>
</tbody>
</table>

29As discussed earlier, the Manchus and Huis have almost completely adopted Han speech.
30Hanitic is the branch of the Sino-Tibetan family with the largest number of speakers by far. The family was called “Indo-Chinese” in earlier writings.

---

6. Zhuang-Dai<sup>31</sup>  
ZHUANG  15,489,630  |  YAY  2,545,059  
BUYI  1,025,128  |  HUAI  20,170,717  
DAM  1,110,900  |  MAONAN  3,091,530  
**31**Groups 6 and 7 are sometimes called Kam-Tai, which is essentially the same group of languages some linguists now call Kachai.

---

8. Austro-Asiatic  
VA  351,974  |  WA  449,716  
BULANG  82,280  |  BENGLONG  15,462  
**31**Groups 6 and 7 are sometimes called Kam-Tai, which is essentially the same group of languages some linguists now call Kachai.

---

9. Turkic  
UYGUR  7,214,431  
KAZAK  1,111,718  
KIRGIZ  1,141,549  
SALAR  87,697  
UZBEK  14,502  
YUGUR  12,997  
TATAR  4,873  
**8,587,067**

10. Mongol  
MONGOL  4,806,849  
DONGXIAN  373,872  
TU  191,024  
DAUR  121,357  
BAOAN  12,212  
**5,505,914**

11. Tungusic  
KOREAN  1,920,597  
XIBE  172,847  
EVENKI  26,312  
OROQEN  6,956  
HEZHI  4,245  
**2,130,969**

12. Austronesian  
BUNUN  “  |  GAOSHANZU  “  
AMI  “  |  “  “  
PUYUMA  “  |  “  “  
PESIWI  “  |  “  “  
SAISIYAT  “  |  “  “  

---

*The Bronze Age and Early Iron Age Peoples of Eastern Central Asia*
investigation of this sort made so far on Chinese populations. The resulting tree diagram, formed on the basis of average linkage, is shown in Figure 3, as reported by Zhao and Lee.

Figure 3: Average linkage tree showing genetic relationships among 74 Chinese populations. [Taken from Zhao and Lee, 1989]

1992. Congruence of genetic and linguistic evolution in China. J of Chinese Linguistics 20:315-331. The average linkage method used for constructing the trees in Figure 1 and Figure 3 implies a constant rate of change across every linkage. For most cases of biological and linguistic phylogeny, there is no reason to believe that this implication holds.

Victor H. Mair, editor
Two points are especially pertinent with respect to Figure 3. One is that the basic dichotomy shown in the tree is once again between northern populations and southern populations. We have seen this dichotomy in Figure 1 above in the discussion of Neolithic peoples. Indeed, Zhao and Lee are able to draw a line on a map of China which separates the two populations quite neatly. This line corresponds approximately to 30 degrees north in latitude. While the exact line must have shifted with population movements over these many millennia, the basic dichotomy between north and south has persisted over a remarkable time span.

The other point is even more pertinent to the present discussion on the relationship between genes and language. For the majority of cases where Hans and non-Hans are sampled at sites closest to each other geographically, they also show the closest relationship in terms of the genetic markers used in the study. In other words, time and again the Hans are closer genetically to their non-Han neighbors than to other Hans who live farther away. Most of the closest relationships shown on the tree, i.e., between 19 and 20, 22 and 23, 26 and 27, 32 and 33, 50 and 51, 62 and 63, are between a Han population and a non-Han population.

Seen in this perspective, genes and language have once again gone their separate ways. Many ethnic groups have preserved their original languages, as well as many other aspects of their culture, even though they have been assimilated into the overall gene pool of the Han majority.

33It is important to keep in mind, of course, that these findings are based on Gmi and Kin allootypes. Other biological traits may yield different results. In an intriguing essay, "Who are the Jews?", Jared Diamond notes that fingerprints may be a more reliable biological marker for deeper affiliations. He notes that "the fingerprints of Ashkenazic Jews still resemble those of their ancient Arab and Egyptian neighbors, rather than those of their recent German neighbors, even though these same Jews have by now become thoroughly Germanized in terms of their ABO blood groups." 1993. *Natural History* 12-19, November.

34This relationship is nicely captured by the saying "Distant relatives are not as important as close neighbors" 译作“远亲不如近邻”, though the saying arose in a different context, of course.

35The following calculation is useful for a preliminary idea of the rate of genetic mixture: “Even a modest trickle of genes can produce great effects if it continues long enough. A classic example is that of African Americans, who today derive an average 30 percent of their gene pool from people of European ancestry. This is the mixture that would have resulted had 5 percent of all black unions been with Europeans in each generation since the institution of American slavery and had all the progeny been classified as black. Another 1000 years of such flow would leave but little of the original African genome.” From p. 78 of L.L. Cavalli-Sforza. 1991. Genes, peoples and languages. *Scientific American* November. A more detailed discussion of the calculations is given on pp. 493-496 of L.L. Cavalli-Sforza and W.F. Bodmer.

36For some discussion of these materials, see footnote 20.


The Bronze Age and Early Iron Age Peoples of Eastern Central Asia

Victor H. Mair, editor

38Three Windows on the Past

39Linguistics

While archeology works with fossils and artifacts, and while genetics looks for traces in the human body, the data for linguistics are old texts and contemporary languages. As we saw earlier, extant texts of the Hans date back at least as early as the oracle bone inscriptions of the Shang dynasty (c. 1200 BCE). Scholars generally share the intuition that the date will recede further back in time as more materials come to light. For example, inscribed pottersdins discovered in Shandong several years ago are believed to be some 4000 years old, although their authenticity and significance are hotly contested.

On the other hand, early texts of the other languages spoken in ancient China are extremely scarce. Furthermore, the analyses of these texts are complicated by the fact that they are usually transcribed in Chinese characters, biased by the Chinese phonology of that time. Nonetheless some progress has been made in this area in recent times.

One example of such investigations is of the Bailangge 百獠歌 of the Han dynasty. This text has been analyzed independently by Coblin, by Ma and Dai, and by Zhengzhang. These scholars conclude that the language recorded in the text represents a Tibetan-Burmese language, most closely related to the Burmese and Yi branches. As such, the Bailangge is especially valuable since it is the earliest sample we have of the Tibetan-Burmese languages; Tibetan was not written down until the 7th century, and Burmese until the 12th century.

Of the characters contained in the Bailangge, Zhengzhang notes that 115 can be related to Burmese, whereas only 40 some characters can be related to Tibetan. From this evidence, he concludes that the two branches of Tibetan-Burmese had already become quite distinct from each other in Han times. He further speculates that at least 2000 years should be added to trace back to Sino-Tibetan unity, which brings it to minimally 4000 years BP.

Another example of progress in the study of ancient minority
Three Windows on the Past

languages is the analysis of the Yueerhe [越人歌] by Wei Qingwen and by Zhengzhang. Here the Han dynasty text is identified as being in a language related to the Zhuang. Yet another area that is being explored has to do with isolated vocabulary items scattered through various classical books. For instance, Li Jingzhong has analyzed many items in the Han dynasty “dialect dictionary”, and proposes relationships between them and Zhuang and other languages of southern China.

Similarly, in chapter 8 of the Manshu[万有], a work compiled by Fan Chuo of the Tang dynasty, there are several dozen items which have been identified with the Bai [白] language of Yunnan. Into this category we can also place the research on cross-language transcriptions in ancient documents, Whether of proper names or of entire phrases. Although such finds cannot always be systematic and numerous, nonetheless they are textual treasures comparable to those uncovered by the archaeologist’s spade. They are riches which have remained largely untapped. Given suitable analysis, they can cumulatively shed much light on the languages of ancient China.

The area in which the bulk of linguistic work has been done is comparative analysis. The famous French Indo-Europeans, Antoine Meillet, has long ago expressed some pessimism regarding historical reconstructions of languages like Chinese. Contrasting with the situation of Indo-European studies, his words were. 11

On n’a presque pas besoin de démontrer qu’une langue est indo-européenne; partout où l’on a trouvé une langue indo-européenne encore inconnue, le ‘tokharian’ ou le ‘hitite’ dans les derniers temps, le caractère indo-européen s’en est révélé dès le début du déchiffrement et de l’interprétation. Au contraire, les langues d’Extrême-Orient qui, comme le chinois ou l’annamite, n’offrent presque pas de particularités morphologiques, n’ont par là même rien qui puisse se prendre au linguiste qui essaie de trouver des langues parentes aux parlars chinois ou aux parlars


By “particularités morphologiques”, Meillet presumably had in mind inflectional systems like verb conjugation and noun declension, judging by his reference to Tocharian and Hititite. It is certainly true that many languages of East Asia and Southeast Asia, in their modern form, are not inflected like Indo-European languages. On the other hand, there is really nothing magical about inflectional systems as a diagnostic for genetic relationship.

Inflectional systems can arise, disappear, be inherited or borrowed, much as any other linguistic trait. Since there is no trait that cannot be borrowed, we have no single litmus test for genetic relationships among languages. The question, rather, is one of probabilities and likelihoods.

The value of an inflectional trait is that it typically involves several words, as in a declensional paradigm, and hence is less amenable to borrowing than single words. In fact, it is likely that several paradigms must be involved before a morphological trait can be transferred from one language to another through contact. While inflectional paradigms are largely absent in many modern languages of East Asia and Southeast Asia, there is reason to believe that they played a more significant role in the grammars of earlier stages of these languages.

On the other hand, there are other types of word families, related not by inflection, but by morphological derivation, compounding, or by semantic connection. These word families may not be as neatly packaged as inflectional paradigms, but their usefulness for genetic investigations should not be overlooked. Indeed, given that Chinese characters often contain semantic information, they provide a source of historical data not available in Indo-European studies.

In fact, this resource was tapped for internal reconstruction in a seminal study by Karlsgen. The approach has been followed up by Pulleyblank, and recently reviewed by Mei.12 Instead of limiting ourselves to an Eurocentric framework based on inflectional systems, we can gain more ground by utilizing the special resources intrinsic to the East and Southeast Asian region, and perhaps arrive at results not achievable elsewhere in the world.

An early hypothesis linking East Asia with the New World is one proposed by Edward Sapir. In a letter written to A. L. Kroeber in 1921, Sapir said:

If the morphological and lexical accord which I find on every hand between Na-Dene and Indo-Chinese is "accidental", then every analogy on God's earth is an accident. ... For a while I resisted the notion. Now I can no longer do so.

Given that the original settlers of the Americas came across the Bering Straits, we have reason to suspect that some languages in Asia are ultimately related to some languages in the Americas. Sapir's contribution lies in his perception that evidence for such a relationship can persist after such long time spans, and in his identification of the Na-Dene group of languages in this relationship, a group of languages which he himself was first to classify.

For various reasons, Sapir did not openly publish much of his work in this area, a topic investigated by Bengston. Recently, however, the idea has received further attention from Starostin, a linguist based in Moscow. The new phylum proposed by Starostin includes not only Sino-Tibetan and Na-Dene, the connection that Sapir worked on, but also four additional clusters of languages with a wide geographical distribution: Basque, Burushaski, Venisenian, as well as some of the languages of the Caucasus.

Basque and Burushaski have long been considered language isolates with relatively small speech communities, the former in northern Spain and the latter in northern India. Venisenian has only a single surviving language, I.e., Ket, spoken in northern Siberia. If this new phylum is indeed a monophyletic unit, the questions that cry for an answer are: when did these languages range over such an immense territory, where did they originate, and how did they acquire such a patchy distribution?

Another genetic hypothesis that has generated considerable interest in recent years is being studied by L. Sagart, a linguist based in Paris. Here the linkage is between Chinese and Austronesian. In many ways, the linkage is at first blush rather surprising. Chinese

---


quality of the data that are used. Unfortunately, the data currently available, for a variety of reasons, are not ideal for the experiment. As ideas from systematics become increasingly applied in linguistics, one would hope that appropriate sets of data for computational purposes will be increasingly available. Nonetheless, the present preliminary effort may be useful toward some tentative conclusions regarding the ancestry of the languages of China.

The data I use for the Chinese dialects are those given in Xu.\textsuperscript{51} The Indo-European data are taken from a recent monograph by Dyen, et al.\textsuperscript{52} For Sino-Tibetan I used the lists published by Benedict for lexical comparisons, with occasional reference to the very useful monograph by Coblin.\textsuperscript{53} Xu's data were compiled for glottochronological calculations. The data of Dyen, et al. were compiled for purposes of lexicostatistic comparisons. Although their purposes are different, the data from Xu and Dyen et al. are readily convertible for use in the present experiment.

The method I used is that first proposed by Fitch and Margoliash, and programmed for the computer by Felsenstein. The results are cross-checked against the neighbor-joining method described by Saitou and Nei, and programmed for the computer by Saitou.\textsuperscript{54} Essentially, the input data are in the form of a matrix which specifies the distance between every pair of languages. In this experiment, this distance is based on the percentage of shared cognates,\textsuperscript{55} as recognized in the three sources.

4. An experiment in subgrouping and dating.

In the remainder of this paper, I will report a preliminary experiment on language grouping, using methods from phylogenetic systematics.\textsuperscript{56} All methods of analysis depend for their success on the

---

\textsuperscript{51}Xu, Tongqiang 徐通锵. 1991. \textit{Liushu yuanyakue 历史语言学}. Shanghai yuanyuguan 上海古籍館. I thank Professor Xu for sending me the unpublished data from which the numbers in his book are derived.

\textsuperscript{52}Dyen, I., J. B. Kruskal and P. Black. 1992. In Indo-European Classification: a Lexicostatistical Experiment. \textit{Transactions of the American Philosophical Society} vol.82, Part 5. I thank Dr. Kruskal for sending me the unpublished data from which the numbers in their monograph are derived.


\textsuperscript{55}The percentage of shared cognates is converted into distance by taking its logarithm.

The Bronze Age and Early Iron Age Peoples of Eastern Central Asia

Victor H. Mair, editor
From this input matrix, the computer program then solves for an unrooted binary tree whose tips correspond to the languages being analyzed. According to the solution, the distance between any pair of languages is represented by the sums of the branch lengths along the shortest path connecting these two languages. Ideally, these distance values will be the same as those specified in the input matrix, though such a perfect match is extremely unlikely, given the complexity of linguistic development. The degree of mismatch between the input matrix and the solved values can be easily calculated. In all three cases under consideration, however, the match is surprisingly good, and none of the solutions need be rejected.

The unrooted tree obtained so far has no directionality; so the next step is to determine the location of the root. A common procedure for doing this is to take the midpoint of the longest path. I have used this procedure in the present experiment, realizing that it may introduce an artificial amount of “leveling” on the evolutionary rates across the lineages. The three trees thus obtained are shown in Figure 4. Note that these trees are significantly unlike those in Figures 1 and 3 in that the branch lengths are typically different from each other.

![Additive trees for Sinitic, Indo-European, and Sino-Tibetan languages.](image)

The Sinitic tree shows that the closest relationship is between Suzhou and Changsha. This is an interesting result because it is paralleled by phonological considerations: the Wu and Xiang dialects are the only ones which have largely preserved the voiced obstruents of Middle Chinese that are lost in all other major dialects. Suzhou, Changsha and Nanchang cluster together as a monophyletic unit, forming a belt of central dialects, as shown in the tree. The southern dialects, on the other hand, show no such clustering.59

The seven Indo-European languages were chosen to maximize the diversity in the group. The deepest split is between the Asian languages and the European ones. It is surprising that the tree shows such great time depth for the separation between Tadjik and Singalese, almost as much from each other as their separation from the European languages.

Another unexpected result is that German groups with Danish rather than English. The received view here, based on phonological development, is that Danish is North Germanic while English and German are West Germanic. Here is another instance, then, where linguistic traits are in conflict—in this case, phonology versus basic vocabulary. Yet, the result from vocabulary should not be surprising, considering the effect centuries of Dane law and Norman rule must have had on the English vocabulary.

The next observation of interest comes from comparing the depth of the Germanic subtree with that of the Sinitic tree directly above it in Figure 4. The fact that these two groups of languages have approximately the same time of divergence—each is about 2000 years old—is nicely captured by the depths of these two trees. This gives us encouragement that perhaps we can use this tree depth as a yardstick to measure the date of divergence of other groups of languages.

Indeed, when we apply this yardstick to the entire Indo-European tree, we find that the tree is roughly three and half times as deep. This gives a divergence date of some 7000 years ago for the family as a whole. At present, there are two competing hypotheses among archeologists regarding this date.60 The “Kurgan invasion” hypothesis, as argued in detail by M. Gimbutas, places the first split at 6000 years ago. C. Renfrew, on the other hand, based on consideration of the diffusion of agriculture, prefers an earlier date of 8500 BP. The date indicated by the trees in Figure 4 is in between these two proposals. At any rate, this result points to the possibility of extending the yardstick to situations which have not received as much attention.

Again, when we apply this yardstick to the Sino-Tibetan tree as a whole, we find that the divergence time is about 3 times as great as that for Sinitic alone. This gives a date of approximately 6000 years.


BP. In a recent study which discusses Sino-Tibetan and Austro-Tai, Peyros and Starostin propose a divergence date of 5000 to 6000 BP. Although they give no justification or procedure for arriving at this number, it is nonetheless consistent with the results arrived at here.

This date may be pushed further back in time, of course, if we should later add to the Sino-Tibetan tree a language whose distance to another language in the family is greater than that between Beijing and Pwo. Similarly, the split dates further back if we later re-classify cognates as borrowings. Conversely, if we are able to increase the percentage of cognacy through deeper phonological relations, the date would move closer in time. Such qualifications apply, of course, to all phylogenetic analysis, and there is no reason to believe that the Sino-Tibetan tree analyzed here is an exception. At the present state of our knowledge, it appears that Sino-Tibetan is a younger family than Indo-European, by perhaps 1000 years or more.

This date of 6000 BP, when the Tibeto-Burman languages first split off from Sinitic, receives some indirect support from the prehistorical scenario that archeologists have constructed. This can be seen from Figure 5, which is adapted from the important synthesis of K. C. Chang.

Figure 5: Linking of cultural spheres at 6000 BP, forming the roots of Chinese civilization. Taken from Chang 1986, p.235.

The three maps in Figure 5 show the geographical distribution of these early cultures at different stages: 9000 BP, 7000 BP, and 6000/5000 BP. The archeological record shows that each cluster of sites expands its domain as its assemblage of cultural artifacts grows more extensive and complex, and, presumably, as the population increases.

The most remarkable fact shown in the map on the right of

---

62A fascinating question is whether these two momentous events in the distant past were related to each other in any way, or whether both were driven by some more global circumstance, created by nature or by man.

---

Three Windows on the Past

Figure 5 is that at around 6000 BP, these cultures began to reveal a significant degree of interaction. Apparently, cultural advancement and population density had reached a threshold by that time. Earlier the clusters were largely independent and isolated from each other, as the map on the left shows for 9000 BP, though they were increasing in scope, as the map in the middle shows for 7000 BP. 6000 BP marked a transformation of these clusters into a linked network, sharing many common archaeological elements. This network was such that the cultural similarities within it were qualitatively greater than with those outside it.

It is reasonable to infer that these similarities were brought about by a significant amount of population movement, so that peoples brought their cultures with them as they moved into new spheres. A linguistic consequence of these movements is the splitting of communities from an original population, such that each community would eventually evolve its own language. In any case, the linkages of these Neolithic cultures were so strong by 6000 BP, that, in the charming words of K. C. Chang, "When the Wei-shui River valley sneezed, as it were, the Lake T'ai-hu region caught cold." He called this network of cultures an "interaction sphere", the "initial China".

Indeed, when the Qin dynasty first unified China in 221 BCE, it was building on this complex amalgam which had begun to come together some 4000 years earlier.

It is appropriate to underscore again the preliminary nature of the experiment reported here, bearing in mind the limitations of the data and the newness of the method as it is applied in linguistics. Our knowledge in this area needs to be significantly deepened before conclusions can be drawn with any degree of confidence. Certainly I am not alone in my hesitation on the use of tree diagrams to represent linguistic history, since these diagrams deliberately ignore the effects of contact. A useful simplification for some purposes, clearly, else these diagrams would not have found such favor in historical linguistics for over a century. Until we eventually can complement the inherited traits with those that are borrowed, our knowledge will remain incomplete and insecure.

Nonetheless, with all due qualifications, the results reported here from the linguistic window appear to be reasonable enough at the present state of our knowledge. It is all the more encouraging, of course, that they seem to correspond in time to events seen through the archeological window. We may hope that with additional and more fine-tuned studies from the genetic perspective, our understanding of the distant past will be enhanced as well. With the coordinated views from these three windows, as well as others yet to be

---

64Chang 1986:410.
explored, we will ultimately be able to derive a truer, multidimensional reconstruction of the past more effectively than any single discipline can hope to offer on its own resources.