

# *Words in the Desert: the Story of Tocharian and the Tocharians*

Morgan Sonderegger

September 26, 2002

There has never been any person dearer to me than you  
and there never will be any dearer.  
The love for you, the delight in you is breath together with life.  
This should not change for life.  
Thus I used to think: with a single beloved will I live well  
lifelong without deceit, without pretence.  
The god Karman alone knew this my thought.  
Therefore, he caused dissension and tore from me the heart that belongs to you.  
He led you away, separated me and had me partake of all sorrows.  
The joy I had in you he took away from me.  
- Tocharian Love Poem

Near the beginning of the twentieth century, the European intellectual community was shocked and amazed by the discovery of texts written in unquestionably Indo-European languages in the distant, exotic Tarim Basin in Chinese Turkestan. Two of these languages, dubbed “Tocharian”, and their speakers, the “Tocharians” have for over a century been the source of both invaluable data and difficult questions for linguists, archaeologists, and historians alike. Research and debate on the history and ancestry of the Tocharian languages and their speakers began when the first Tocharian text was brought to the West, and has hardly stopped since. Why have the questions associated with these languages captivated so many for so long? Perhaps because of the unique nature of what is known about the Tocharian languages and their speakers. Of the thousands of pages of documents in Tocharian that make up the only artifacts known to be left behind by the Tocharians, almost all are Buddhist texts that offer little insight into the culture or nature of this mysterious people. As the Tocharians would have lived in towns along the diverse Silk Road, geography alone cannot tell us very much about their precise identity. In short, the only sure source of

information about the speakers of Tocharian (aside from a few wall paintings) is Tocharian itself - a relatively unique situation among the ancient peoples of Eurasia that we know of<sup>1</sup>. Nonetheless, work on the Tocharian languages and the identity of the Tocharians has slowly progressed through the efforts of academics of many stripes, and although many aspects of the so-called “Tocharian problems” remain matters of speculation, a look at what has been accomplished thus far is interesting and rewarding. This paper presents a brief summary of how the Tocharian languages were discovered, their role (linguistic and otherwise) in Indo-European studies, and what is known about them and their speakers. Although some space is also devoted to the many linguistic, historical, and archaeological speculations surrounding the exact origins and ancestry of the Tocharian languages and their speakers, a full-length book could barely do justice to the sprawling body of literature on the subject, and we do not discuss it at length here.

Beginning in the late nineteenth century, expeditions to the Tarim Basin region of what is now the Chinese province Xinjiang found evidence of many cultures previously unknown to the West. Among the artifacts found during these expeditions were writings in two hitherto unknown languages (although it was not immediately recognized that they were two different languages, since they utilized the same writing system) which we will call A and B for the moment. The Moravian missionary Weber procured the first documents in A and B from local peoples who had found them near Jarkand and in Kucha in the 1890s. Over the next few decades, Russian, English, Japanese, French, and German expeditions led by explorers such as Aurel Stein, Albert Grünwedel, Albert von Le Coq, and Paul Pelliot found many more texts in these two languages, primarily in B<sup>2</sup>. Discovery of texts in A and B continued at a lesser rate throughout the twentieth century, with the most recent major discovery in 1974. In total, 3,640 pages or fragments of text are known, comprising about 4,500-5,000 words of A and B.<sup>3</sup> Several other languages were discovered in the course of the aforementioned

---

<sup>1</sup>This is not the case for the entire world, though - some Native American peoples are known only through their languages, recorded by missionaries who translated the Bible into the local vernacular before the extinction of its speakers (thanks to Norvin Richards for pointing this out).

<sup>2</sup>Van Windekens, pp. xiii-xiv

<sup>3</sup>Mallory & Mair, pp. 270-271

expeditions, and have played an important role in Tocharian studies due to their linguistic influence on the Tocharian languages. Documents in a language dubbed Kroranian were found in the Southern Tarim Basin, while documents in Iranian languages such as Saka and Sogdian were found in areas to the north, west, and south of the Tarim Basin.

The languages A and B were quickly recognized as different but closely related upon study in the West. Their fundamentally Indo-European nature was recognized by F.W.K. Müller and A. von Le Coq in 1907, and they were given the names Tocharian A and Tocharian B by E. Sieg and W. Siegling in 1908<sup>4</sup>. Sieg and Sieling based this naming on a colophon to an Uighur translation of a Buddhist text, the *Maitreyasamiti*<sup>5</sup>, stating that the text had not been translated directly from Sanskrit, but from another language<sup>6</sup>:

“The sacred book Maitreya-samiti which the Bodhisattva guru ācārya Āryacandra, who was born in the country of Nagaradeśha, had composed in the *toχri* language out of the Indian language, and which the guru ācārya Prajñaraksita, who was born in Il-bliq, translated from the *toχri* language into the Turkish language.”

When Sieg and Siegling read this colophon, the historical Tokharoi people, whose Kushan empire was relatively close to the Tarim Basin in the first few centuries AD, was already well known to Europeans from historical sources in several languages (Sanskrit: *Tukhāra*, Greek: *τοχαροι*, Latin: *Tochari*, Chinese: *Tuholo*). Based on the phonetic similarity between the name “Tokharoi” in these ancient languages and the name of the *toχri* language, combined with the fact that the only known translations of the *Maitreyasamiti* are in Uighur and the A language, Sieg and Siegling concluded in 1908 that the language A was the same as the *toχri* language that the *Maitreyasamiti* was translated from and that the *toχri* language was the language of the Tokharoi, and named A and B Tocharian A and Tocharian B<sup>7</sup>. The names have stuck, but this naming was probably not justified, as discussed later on. From

---

<sup>4</sup>Adams, p. 1

<sup>5</sup>Translated by F.W.K. Müller

<sup>6</sup>Ibid., p. 2

<sup>7</sup>Ibid.

here on, the speakers of Tocharian languages will be referred to as “Tocharians”, while the people of the Kushan empire will be referred to as the “Tokharoi” or “Tokharians”.

No Rosetta stone was required to decipher the Tocharian languages, which were written in a script closely related to the Brāhmi scripts (such as devanāgarī). Furthermore, almost all the Tocharian writings found corresponded to known Buddhist texts in other languages (mostly Sanskrit and Prākrit, but also Chinese, Turkic languages, Tibetan, Chinese, and Saka) <sup>8</sup> - some texts are even bilingual versions. The Tocharian texts that have been found are mostly Buddhist writings of various sizes, often single leaves of religious texts left as offerings at desert shrines . They were then covered by sand and preserved in almost perfect condition until their discovery<sup>9</sup>. Other than religious texts, wooden caravan tablets and small numbers of documents on grammar, astronomy, history, medicine, and business affairs have also been found <sup>10</sup>. All samples of written Tocharian A and B that have been found date from the 6th to the 8th century AD.

Tocharian A and B were quickly deciphered and analyzed by Western linguists. They turned out to be unequivocally Indo-European, which was not altogether surprising in itself – they were surrounded by the Iranian and Indic (both Indo-European families belonging to the Indo-Iranian family) languages mentioned above, had many loanwords from these languages, and used a derivative of a writing system employed by Indic languages. But in terms of linguistic structure, the Tocharian languages are no more like the Indo-Iranian languages than any other Indo-European language. As J.P. Mallory puts it, although there are a plethora of hypotheses on the relation of the Tocharian languages to other Indo-European languages, “Those who have attempted to associate Tocharian with other Indo-European language stocks appear to be in agreement that there are no grounds whatsoever for seeing a special genetic relationship with Indo-Iranian”<sup>11</sup>, meaning that the Tocharian languages are more closely related to the far off Indo-European languages of the West than to their

---

<sup>8</sup>Mallory & Mair, p. 272

<sup>9</sup>Renfrew, p. 65

<sup>10</sup>Mallory & Mair, p. 272

<sup>11</sup>Ibid., p. 292

nearest Indo-European neighbors!

Since the discovery of the Indo-European character of Tocharian, linguists have pored over Tocharian texts, and have at times claimed genetic relationships between Tocharian and the Germanic, Balto-Slavic, Italic, Greek, Celtic, Phrygian, or Anatolian languages<sup>12</sup>. While some of these claimed relationships (especially Germanic) seem to have more support than others, that the above families make up the entire Indo-European language family save the Armenian and Tocharian languages indicates that there is no real consensus as to which subfamilies of Indo-European the Tocharian subfamily is related to. Tocharian is usually designated as its own subfamily in family trees of the Indo-European languages, related to the other Indo-European languages only by a common progenitor, Proto Indo-European (PIE). In this context, it seems prudent to talk about why there is a solid consensus that Tocharian is an Indo-European language, even though its position within the Indo-European family is ambiguous.

Basic techniques of historical linguistics applied to the Tocharian languages reveal their Indo-European character. For example, languages that are related often share many cognates, which is certainly the case with Tocharian and the other Indo-European languages. The table below shows some cognates (and the two non-cognate words “man” and “right”) shared between Tocharian and various other Indo-European languages [10]:

---

<sup>12</sup>Ibid., p. 286-287

English	Toch. A	Toch. B	Old High German	Greek	Latin	Sanskrit
one	sas	še	ein	hen	unus	éka
two	wu	wi	zwā	duo	duo	dvi
three	tre	trai	drī	treis	tres	trī
four	štwar	štwer	fior	tettares	quattuor	catúr
five	pän	piš	fimf	pente	quinque	pañcan
six	šäk	škas	sēhs	hex	sex	ṣaṣ
seven	špät	šukt	sibun	hepta	septem	saptán
eight	okät	okt	ahto	okto	octo	aṣṭan
nine	ñu	ñu	niun	ennea	novem	návan
ten	šäk	šak	zehan	deka	decem	dashan
father	pācar	pācer	fater	pater	pater	pītā
mother	mācar	mācer	muoter	meter	mater	mātar
brother	pracar	procer	bruoder	phrater	frater	bhrātā
sister	šar	šer	swestar	adelfi	soror	bhaginī
man	napem	šauomo	mann	andros	homo	manushhya
right	pāccās	saiwai	reht	dexia	ius	dakhiña

The most significant fact about the cognates shared by Tocharian A and B and other Indo-European languages is the similarity of the basic numerals in Tocharian to those in PIE and other Indo-European languages, since most Indo-European languages have basic numeral cognates with PIE and with each other. Additional clues that the Tocharian languages are Indo-European lies in their phonology<sup>13</sup>, which can be reconstructed from the syllabic script Tocharian A and B are written in. The phonological system of Tocharian is, at the phonetic level, rather different from its parent, PIE. However, the phonological correspondences (how each sound in the parent language mapped to its corresponding sound in the daughter language through language change) between PIE and Tocharian have been worked out, and predict quite well the sound<sup>14</sup> of PIE cognates in Tocharian. That such a clean correspondence can be established is strong evidence for a relationship between Tocharian and PIE. That verbs and nouns in Tocharian seem to act like an evolved version of their PIE equivalents is further evidence. Tocharian also has some linguistic similarities to Indo-European

<sup>13</sup>The sound system and phonemes of a language.

<sup>14</sup>The orthography, technically, since no living speaker of Tocharian exists to confirm that the orthography of Tocharian A and B reflects the actual sound patterns of these languages. However, the orthography of languages written using syllabic writing systems, such as Tocharian A and B, usually reflects fairly accurately the phonology of these languages, due to the phonetic nature of the writing system, which allows accurate transcription of spoken language.

languages other than PIE. For example Tocharian, like the Germanic languages, has many n-stems <sup>15</sup>. Many other correspondences between Tocharian and the Indo-European languages leave little doubt that Tocharian is itself an Indo-European language <sup>16</sup>.

Due to their exotic location and structure, the Tocharian languages have played an important role in Indo-European linguistics since their discovery. Any theory of the origin and spread of the Indo-European languages must account for the geographical position and linguistic structure of the Tocharian languages, a daunting criterion that helped force the reevaluation of some established ideas in Indo-European linguistics at the time of Tocharian's discovery.

For example, classical Indo-European linguistics holds that the each Indo-European language can be put into one of two groups based on what sound the PIE *\*k̑* has become in that language. Languages in which this sound stays in the back of the mouth (as /k/, /g/, /h/, etc.) are called *centum* languages after the Latin word starting with /k/ which corresponds to the PIE word for “hundred” ( *\*k̑mtom* ). Languages in which this sound becomes a sibilant (like /s/ or /ʃ/ ) are called *satem* languages, after the Persian word *satem* corresponding to *\*k̑mtom*. Prior to the discovery of the Tocharian languages (and the ancient Anatolian language Hittite), the Indo-European languages could be neatly divided by the so-called *centum/satem* distinction, roughly signifying western and eastern languages. Linguists imagined that speakers of PIE at some point split into two groups which then headed in opposite directions, one developing each type of language. The Tocharian languages largely undermined confidence in this distinction, as they are *centum* languages (the Tocharian A and B words for “hundred” are *känt* and *kante*) which are found in *satem* territory (very far east of Europe). If such a simple split is to be postulated, the ancestors of the speakers of Tocharian must have somehow headed west with the *centum* languages to Europe, then turned around and trekked several thousand miles to the east and eventually ended up in the Tarim Basin. Given this unlikely scenario (and extra evidence from Hittite), most historical

---

<sup>15</sup>Meaning roughly that a broad range of suffixes end in n.

<sup>16</sup>Mallory & Adams, p. 591-592

linguists agree that the *centum*/*satem* distinction should not be treated as synonymous with a West/East one<sup>17</sup>.

Given the difficulty of explaining the relationship of Tocharian to the other Indo-European languages with a family tree model of the Indo-European languages which strictly defines each subfamily's relationship with the others, Tocharian also helped inspire more modern models of the relationships among the Indo-European languages, such as the “wave model” of Johannes Schmidt<sup>18</sup>, in which linguistic changes propagate (like waves) throughout nearby languages, and a language family is divided up into its component subfamilies by lines indicating linguistic differences. One hypothesis for the relative position of the Tocharian languages which can be pursued under such a model of the Indo-European languages holds that Tocharian, Germanic, Italic, and Hittite are “periphery languages” whose speakers separated from the rest of Indo-European language speakers relatively early on<sup>19</sup>. Indeed, if there is any regularity in the distribution of *centum* languages such as these, it is that they are on the geographic periphery of where Indo-European languages are known to have been spoken. In this model, the /k/→sibilant change in the *satem* languages occurred after speakers of the *centum* languages) had left the prehistoric center of the Indo-European world.

Having seen the unequivocally Indo-European nature of Tocharian A and B, we now turn to what is known about them and their relationship to each other.

Tocharian B texts have been found across the northern Tarim Basin along what was the northern branch of the Silk Road. Many texts have been found in the territory of the ancient state of Kucha and in the surrounding caravan passes. Today, most specialists agree with Sylvain Lévi, who in 1913 analyzed the Tocharian B documents found in the Kucha area and proposed that Tocharian B was the language of the Kuchean state<sup>20</sup>. Hence, Tocharian B is sometimes called Kuchean. Tocharian A texts have been found only in the Eastern part of the region where Tocharian B texts have been found, mostly near Turfan

---

<sup>17</sup>Mallory, p. 154

<sup>18</sup>Renfrew, p. 105

<sup>19</sup>Mallory, p. 61

<sup>20</sup>Adams, p. 6



and Qarashähär, a town which was once part of the Agni kingdom. Hence, Tocharian A is sometimes called Turfanian, Karashahrian, or Agnean. Texts in Tocharian B are much more common than those in Tocharian A even in areas where both are found. All non-religious Tocharian writings that have been found are in Tocharian B, which appears to have been the vernacular in the areas where these texts were found. There is a large body of evidence suggesting that when the Tocharian A and B texts were written, Tocharian A was a “dead” liturgical language, while Tocharian B was the “living” vernacular. For example, in some Tocharian A texts, particularly difficult words are glossed in Tocharian B, suggesting that the copier of the Tocharian A text wanted to remind himself in the vernacular of these words’ meaning. The same situation obtains in European medieval religious texts, in which monks would sometimes scribble in the vernacular definitions of difficult words in Latin, the liturgical language. Also, the one purely secular piece of Tocharian literature that has been found is a love poem, written in Tocharian B (reproduced at the beginning of this paper)<sup>21</sup>. This liturgical/vernacular distinction would help explain the fact that Tocharian A and B are markedly distinct languages, even though they were apparently used within a few hundred miles of each other. The linguistic difference between Tocharian A and B has been likened by Douglas Adams to that between Italian and Romanian<sup>22</sup> - two languages which are closely related, but whose speakers could probably not understand each other. A broad linguistic estimate of the time two groups of speakers of a “Proto-Tocharian” would need to be separated before the languages they spoke would differ as much as Tocharian A and B is 500-1000 years<sup>23</sup>. One mystery of the Tocharian languages is how Tocharian A and B could have become so different when their respective speakers (if they lived in the same sites as the texts found in their languages) lived a few hundred miles apart on the well-traveled Silk Road.

Tocharian A and B present us with another mystery: linguistically, Tocharian B should be the “dead” language and Tocharian A the “living” one. In terms of linguistic structure,

---

<sup>21</sup>Mallory & Mair, pp. 273-277

<sup>22</sup>Ibid., p. 275

<sup>23</sup>Ibid., p. 276

Tocharian B is markedly more conservative than Tocharian A in that it seems to have preserved much more of the structure of Proto-Tocharian, the hypothesized parent language of the two. For example, words in Tocharian A often have a syllable or ending dropped compared to their Tocharian B cognates. Under normal circumstances, this could serve as evidence that Tocharian B is older than Tocharian A, yet we know this is not the case<sup>24</sup>.

The Kuchean, the hypothesized speakers of Tocharian B, are the only group of people generally accepted to have spoken Tocharian languages. Since the Tocharian writings we have tell us little about the actual identity of the speakers of Tocharian beyond that Tocharian B was most likely the language of Kucha, it is quite difficult to say anything about the speakers of Tocharian based solely on the location of the Tocharian texts that have been found. Since the region where Tocharian B writings are found was part of the cosmopolitan Silk Road, non-Kuchean peoples in the area may well have spoken Tocharian languages, and the Kuchean themselves may have spoken other, non-Tocharian languages. In addition, given the religious nature of most texts that have been found in Tocharian A and B, all we know for sure (excepting the case of Kucha) is that Tocharian was relatively widely used for religious (Buddhist) purposes. Both Tocharian A and B could have simply been liturgical languages for the non-Kuchean, a view that has been advanced in differing forms by George Lane and Werner Winter<sup>25</sup>. In this case, we would know little about the people who spoke the Tocharian languages besides that some of them lived in Kucha, a disappointing prospect.

Some hope comes from excavations of the ancient kingdom of Kroraina along the southern branch of the Silk Road in the southern Tarim Basin. Administrative documents written in the Indic language Prākṛit have been excavated from sites in the territory of ancient Kroraina<sup>26</sup>. These documents contain several hundred clear loanwords, mostly place names and personal names, which have been found to be neither of Iranian or Indic origin. Since they refer to local places or people, these words are assumed to have been borrowed from the native language of at least some portion of the Kroranian population. In the 1930s,

---

<sup>24</sup>Ibid., p. 277

<sup>25</sup>Adams, p. 6

<sup>26</sup>Barber, pp. 117-118

the linguist Thomas Burrow analyzed these words and found that they were probably of Tocharian origin, meaning that they came from some language in the Tocharian group of languages. Very few of the words are actual cognates of words in Tocharian A or B, but their phonetic and morphological structure closely reflect the phonology and morphology of the Tocharian languages of the northern Tarim Basin <sup>27</sup>. These words are postulated to come from a third, southern dialect of Tocharian, dubbed Tocharian C and sometimes referred to as Kroranian.

Tocharian C holds tantalizing clues to the identity of the speakers of the Tocharian languages. The Kroranian texts were found in the southern Tarim Basin, where no Tocharian A or B texts have been found, and appear to date from the third century AD, several hundred years before the appearance of (written) Tocharian in the northern Tarim Basin. This makes it very likely that there was a non-Kuchean people in the southern Tarim basin who spoke a Tocharian language. It has been proposed that the speakers of Tocharian C were converted to Buddhism as it made its way along the southern Silk Road in the early part of the first millenium AD<sup>28</sup>. This people would then have either made their way to the northern Tarim Basin over the next 300 years, or both northern and southern groups of Tocharian speakers may have existed. Although postulating such a broader distribution and migration of Tocharians is somewhat of a leap of faith given the lack of direct evidence, that there is no evidence (so far) of a Tocharian-speaking people other than around the oasis towns of the Tarim Basin is hardly surprising, given the harshness of the terrain. As Colin Renfrew puts it, “It is only because there were oasis cities in this vast area of steppe land which had an urban and literate culture that we have any knowledge whatsoever of this long-forgotten language [Tocharian]. It is very possible... that there were other groups in the intervening territories speaking Indo-European [Tocharian] languages which have less evident traces.”<sup>29</sup>. Keeping Renfrew’s observation in mind, we now turn to the controversial topic of the origins of the Tocharians, the most heavily debated issue in Tocharian studies.

---

<sup>27</sup>Mallory & Mair, pp. 277-279

<sup>28</sup>Mallory & Mair, p. 279

<sup>29</sup>Renfrew, p. 67

Interest in the Tocharians seems to have stayed robust for over a century largely because there is strong evidence that they were a Europoid people speaking a language related to the languages of Europe, yet living on the edge of Han China. As noted above, Tocharian is an Indo-European language which appears to be most closely related to the ancient Indo-European languages of the West. Chinese Han Dynasty sources indicate that of the peoples (which they were careful to differentiate from one another) surrounding the area where the Tocharians would have lived, the Wu-Sun had red hair, blue eyes, and “resembled monkeys”<sup>30</sup>, and the Yuezhi were hairy and had white skin <sup>31</sup>. Most striking of all, wall murals that later Tocharians drew of themselves donating to Buddhist temples show men with moustaches, high-bridged noses, and brown and red hair <sup>32</sup>. Given this evidence that the speakers of Tocharian were physically Caucasian, the burning question is how and when their ancestors originally arrived in Xinjiang. Although this question has been a matter of hot debate and has solicited many theories of Tocharian origins for over a century, none of these theories is generally accepted. Part of what makes this “Tocharian problem” so hard to solve is that since the Tocharians were an Indo-European people who split off from the PIE stock early on, a comprehensive theory of Tocharian origins must not only explain the sparse evidence we have of the Tocharians, it also must be consistent with a theory of the earliest Indo-European peoples and their gradual spread. Since the question of where the Indo-European homeland was and how the Indo-Europeans dispersed from it is anything but decided, a solution to this Tocharian problem is made that much harder.

Another major Tocharian problem is whether or not the name “Tocharian” itself is an appropriate name for the Tocharian languages. We saw earlier how Sieg and Siegling equated the language we now call Tocharian A with the language of the historical Tokharoi people, whose existence and history is recorded in written sources in several languages. The Tokharoi are generally accepted to be the same as the Yuezhi people often referred to in Chinese sources. According to Han dynasty records, the Yuezhi originally lived directly west of the

---

<sup>30</sup>Mallory, p. 60

<sup>31</sup>Barber, pp. 118-119

<sup>32</sup>Ibid., p. 118

Gansu corridor in what is now China, but were defeated by the Xiongnu sometime around 176 BC. Rather than stay as vassals of the Xiongnu, the Yuezhi left their homeland<sup>33</sup>. They eventually split into two groups. The “Greater Yuezhi” eventually defeated the historically Greek residents of Bactria and founded the Kushan empire, while the “Lesser Yuezhi” settled among the Qiang people, probably in the mountains directly south of the Tarim Basin <sup>34</sup>. The Yuezhi were almost certainly a Europoid or Causacoid people, and thus probably spoke an Indo-European language. The main question is whether the Yuezhi/Tokharoi (referred to as the Yuezhi from here on to avoid Tokharoi-Tocharian confusion) people are ethnic and linguistic siblings of the Tocharians whose texts we find in the Tarim Basin. Much evidence for and against such a relationship has been proposed, but a safe conclusion seems to be “we don’t know, but probably not.” The main argument for a Yuezhi-Tocharian relationship stems from the Uighur colophon discussed above. Sieg and Sigling’s judgment that the *toχri* language of this colophon is the same as the language of the Tokharoi and that this language is Tocharian A has been assualted on many fronts. Some have pointed out that this colophon shows only what the Uighurs called the *toχri* language, not what its speakers called it. Also, there is no definitive evidence that *toχri* was the language of the Tokharoi to begin with. The other major pieces of evidence for a Tokharoi-Tocharian relationship are one Tocharian B text that glosses “Kuchean” with “Tokharian”, and another in which the Sanskrit words for “a Tokharian woman” and two Tocharian B words interpreted to be “a Kuchean woman” follow each other. However, linguistic analysis of the second text has shown serious problems with the proposed translation (even insofar as the Tocharian B words refer to anything “Kuchean” at all), and the first text is not solid evidence. The term “Tokharian” seems to have been used broadly at the time the first text was written, and offers no assurance that “Tokharian” actually refers to the historical Tokharoi we are considering<sup>35</sup>. Although there are many further arguments for and against a definite relationship between the Yuezhi and the Tocharians, we move on, noting that most Tocharian specialists seem to not put stock

---

<sup>33</sup>Barfield, p. 39

<sup>34</sup>Mallory & Mair, pp. 58-59

<sup>35</sup>Mallory & Mair, pp. 333-334

in such a relationship.

Whoever the Tocharians of the first millenium AD actually were, the more basic question is how their ancestors got to the Tarim Basin to begin with. Two books ([6], [2]) deal with this controversial question in great detail in the context of the identity of the mysterious Tarim Mummies<sup>36</sup>, and the reader is referred to them for a thorough (if sometimes biased) discussion of theories of the identity of the ancestors of the Tocharians. To summarize briefly, the theories presented in these two books interpret somewhat differently the evidence found in the tombs of the Tarim Mummies. Elizabeth Barber, a linguist and textile specialist, notes uncanny similarities between the textiles of the Tarim Mummies and the textiles of certain ancient Celtic cultures of Europe. Barber ties the Tocharians to the Yuezhi, explores the similarity of the Tocharian and Celtic language families, and hypothesizes that the ancestors of the Tocharians originally came from Europe, and at some point made the long trek east. In their book, James Mallory, an archaeologist, and Victor Mair, an archaeologist and historian, tentatively support a hypothesis in which the ancestors of the Tocharians were members of the Europoid Afanasevo culture which inhabited Southern Siberia about 4,000-5,000 years ago. These proto-Tocharians would then have headed south, penetrating the Tarim Basin from the east and possibly founding the Qäwrighul culture of c. 2000 BC of which some of the Tarim Mummies were members. Mallory & Mair base their argument on extensive consideration of the linguistic, historical, and archaeological record.

In the end, it must be admitted that any grand solution to the many questions surrounding the Tocharians and their languages is still quite tentative. However, Tocharian studies has advanced greatly since its inception, and new archaeological techniques and finds combined with a score of ambitious, cross-diciplinary researchers boasting expertise in a variety of social and physical sciences are bringing the “Tocharian problems” ever closer to resolution. When these problems are solved, we will have learned not only the story of the Tocharians and their languages, but a vast amount about the cultures, peoples, and languages of the

---

<sup>36</sup>Mummies with apparently Causacoid features which have been found in the Tarim Basin. The origin, language, and date of whatever culture(s) these mummies came from is the subject of much research and debate. A good short introduction to their discovery and significance is [?].

Tarim Basin and Central Asia. As happens so often in the study of Central Asia, the story of a people and their language has turned out to be intertwined with the stories of other peoples, places, and languages - in this case spanning almost the entire Eurasian continent. The study of Tocharian and the Tocharians will surely hold new surprises for years to come as scholars slowly close in on their identity and ancestry.

## References

- [1] Adams, Douglas Q. (1988). *Tocharian Historical Phonology and Morphology*. New Haven: American Oriental Society.
- [2] Barber, Elizabeth W. (1999). *The Mummies of Ürümchi*. New York: W. W. Norton.
- [3] Barfield, Thomas J. (1989). *The Perilous Frontier*. Cambridge: Blackwell.
- [4] Mair, Victor (1995). “Mummies of the Tarim Basin,” *Archaeology* 48 (2): 28-35.
- [5] J.P. Mallory & D.Q. Adams, editors (1997). *The Encyclopedia of Indo-European Culture: “Afanasevo Culture”, “Andronovo Culture”, “Indo-European Homeland”, “Qäwrighul Culture”, “Tocharian Languages”*. London: Fitzroy Dearborn.
- [6] Mallory, J.P. & Mair, Victor H. (2000). *The Tarim Mummies*. London: Thames & Hudson.
- [7] Mallory, J.P. (1989). *In Search of the Indo-Europeans*. London: Thames & Hudson.
- [8] Renfrew, Colin (1987). *Archaeology and Language*. London: Jonathan Cape.
- [9] Van Windekens, A.J. (1944). *Lexique Etymologique des Dialectes Tokhariens*. Louvain: Bureaux du Museon.
- [10] The table of words in different Indo-European languages was put together from words found in [6], [2], and various dictionaries.