

# THE POSITION OF TOCHARIAN AMONG THE OTHER INDO-EUROPEAN LANGUAGES

DOUGLAS Q. ADAMS

UNIVERSITY OF IDAHO

The placement of Tocharian vis-à-vis the other Indo-European groups has been something of a problem; those who have investigated the issue over the past seventy-five years have, on the basis of very limited evidence, come to a variety of conclusions. An investigation here of the phonological, morphological, and lexical isoglosses (and especially the latter) uniting Tocharian with other Indo-European language families leads to the grouping of Tocharian with Meillet's "Northwestern" group in the first instance, with particularly close ties with Germanic. The ties Tocharian has with both Greek and Indic outside of the Northwestern group reflect later contacts on the part of the pre-Tocharians with the pre-Greeks and pre-Indics as the pre-Tocharians moved progressively eastward in the late Proto-Indo-European world.

1. Since their identification as Indo-European languages in the early years of this century, the exact relationship of Tocharian A and B with the other Indo-European groups has been the source of considerable debate, particularly since Tocharian did not seem to fit neatly into any of the received categories of Indo-European dialectology. Indeed, its early recognition as a *centum* language (Tocharian A *känt*, B *kante* 'hundred') in the "far east" of Indo-European languages where only *satem* languages were expected robbed that isogloss of its once primary position. However its positive contribution to our knowledge of Indo-European relationships has been more elusive, as the following list of often conflicting hypotheses advanced by various investigators shows:

- 1913 Related to Italo-Celtic (H. Pedersen)
- 1914 Between Italo-Celtic on the one hand and Slavic and Armenian on the other (A. Meillet)
- 1917 Related most closely to Celtic (J. Charpentier)
- 1922 Related to Phrygian and, more distantly, to Armenian (E. Hermann)
- 1923 Related in the first instance to Thracian and Phrygian, in the second to Armenian and Balto-Slavic, and in the third to Germanic (J. Pokorny)
- 1925 Part of a group with Italo-Celtic, Phrygian, and Hittite (H. Pedersen [this proposal was repeated in 1938 and 1941])
- 1936 Between Italo-Celtic on the one hand and Greco-Armenian on the other (E. Benveniste)
- 1954 Originally most closely connected with Balto-Slavic and Germanic, more distantly with Greek and Thracio-Phrygian (W. Porzig)

- 1970 Between Balto-Slavic and Greco-Armenian (a la Benveniste) with later contacts with Italo-Celtic and Slavic (with whom they enjoyed their last Indo-European contacts) and NO particular relationship with Germanic (G. S. Lane)

Until the late thirties Tocharian was not well enough known to provide meaningful data about its relationship with other Indo-European groups, so it is only with Holger Pedersen that any of the proposed connections takes on solidity. He based his conclusion very largely on Tocharian's possession of the "r-passive" (e.g., *kaustär* 'he is killed'). It was his opinion that the "r-passives" in Indo-European represented an innovation and thus Tocharian was to be grouped with Italo-Celtic, Phrygian, and Hittite since they all shared this particular innovation. Subsequent investigators, however, have seen the "r-passive" as an archaism and thus its presence is not diagnostic of any particularly close relationship among groups that retained it.

While showing differences among themselves, there is a certain commonality in the conclusions reached by Benveniste, Porzig, and Lane. All agree that there is some connection of Tocharian with Balto-Slavic and Greek. Beyond these groups there may be some connection with Germanic (Porzig [but specifically denied by Lane]), Italo-Celtic (Lane), or Thracio-Phrygian (Porzig). There is also unanimity in not seeing any special relationship with Indo-Iranian. All of these investigators, including Pedersen, used what we might call "traditional" methodology—grouping the various branches of Indo-European on the basis of those characteristics which intuitively were diagnostic by

virtue of their being shared innovations. Of the three later investigators, only Porzig made explicit his reasons for choosing one isogloss rather than another.

2. In the thirties Kroeber and Chrétien (1937 [see also Kroeber, 1960]) experimented with a more statistical and perhaps less subjective technique in which they took a list of variables gleaned from Meillet's *Les Dialectes indo-européens* and established their presence or absence in each of nine Indo-European groups (Albanian, Tocharian, and initially Hittite were excluded for lack of data) and then calculated, by means of formulas, the degree of resemblance of one group with another. Their method does not attempt to distinguish common retentions from common innovations but it does purport to show the overall phonological and morphological closeness of any pair of Indo-European groups (no lexical data was involved). Since their results agreed very closely with classifications based on more traditional means, it seems reasonable to assume that their methodology really does measure what it is supposed to.

I have supplied the data for Tocharian and calculated the indices on the basis of Kroeber and Chrétien's formula  $Q_2$  ( $Q_2 = (ad - bc) / (ad + bc)$  where  $a$  is the number of shared positives,  $b$  the number of positives in one language matching negatives in the other,  $c$  the number of negatives in one language matching positives in the other, and  $d$  the number of shared negatives). The results appear below (and should be compared to their Table IV [1937: 99]).

TABLE I  
KROEBER-CHRÉTIEN, 74 ELEMENTS, FORMULA  $Q_2$   
(Tocharian Added)

	Ce	La	Gr	Ar	Ir	Sk	Sl	Ba	Ge	To
Ce	1.00	.94	-.37	-.67	-.41	-.42	-.26	-.42	-.17	.12
La	.94	1.00	.13	-.66	-.81	-.82	-.54	-.47	.15	.24
Gr	-.37	.13	1.00	.35	.30	.39	-.33	-.23	-.25	.34
Ar	-.67	-.66	.35	1.00	.29	.07	.42	.38	-.27	.30
Ir	-.41	-.81	.30	.29	1.00	.96	.42	.15	-.31	-.46
Sk	-.42	-.82	.39	.07	.96	1.00	.26	.17	-.75	-.24
Sl	-.26	-.54	-.33	.42	.42	.26	1.00	.96	-.15	.13
Ba	-.42	-.47	-.23	.38	.15	.17	.96	1.00	.43	.17
Ge	-.17	.15	-.25	-.27	-.31	-.75	-.15	.43	1.00	.27
To	.12	.24	.34	.30	-.46	-.24	.13	.17	.27	1.00

As this table suggests, the statistical methodology used by Kroeber and Chrétien tends to confirm the traditional conclusions of Benveniste, Porzig, and Lane. We find Tocharian most like Greek and Armenian

(though the degree of resemblance between Armenian and Balto-Slavic that the table suggests seems too high and I think the same is true of the degree of resemblance suggested to exist between Armenian and Tocharian) on the one hand and like Germanic and Italic (and to a lesser extent like Baltic) on the other. The distance of the relationship with Indo-Iranian is confirmed. (Indeed, only the [negative] index of resemblance with Iranian is statistically significant at the .05 level of confidence. The other indices, while not significant at that level, are suggestive.)

It is also instructive to look at the Tocharian correlations from the point of view of the other Indo-European groups. We find for instance that, aside from Baltic, Tocharian ranks as the group most nearly similar to Germanic. Similarly for both Italic and Celtic, Tocharian ranks in second place after Celtic and Italic respectively. These data (admittedly statistically questionable) indicate a closeness of relationship with Meillet's "Northwestern Group" that has not been commented on before.<sup>1</sup>

<sup>1</sup> Considering the general paucity of our knowledge of agricultural vocabulary in Tocharian, it is perhaps significant that a certain number of those mostly agricultural words Meillet singled out as characterizing his Northwest Group are found in one or the other or both of the Tocharian languages: \*bedh- 'dig' in Tocharian A *pate* 'plow' (\*bhodho-) and *pātā* 'to plow' (\*bhodh-eA-);

\*A(e)w(e)sa- 'oats' in Tocharian B *ysāre* 'wheat (vel sim.)' [A *wsār*] (the *Klosterrechnungen* demonstrate conclusively that the Tocharian B word *ysāre* does not mean 'grain/corn' in general [which is *šātre*]; since Kucha was known to Chinese travelers [cf. Lévi, 1913: 353] for its wheat and millet [the latter probably *yap*], I take *ysāre* to be 'wheat,' the other named grain in our texts from a PIE \*Awesā-ro- [cf. Latin *avēna*]);

\*k̑rAsn- 'hornet' in Tocharian B *kronkše* 'bee' from a PIE \*k̑rAsnukuken- (?);

\*keAw(d)- 'strike' in Tocharian B *kau-* [A *ko-*] 'to kill' and *kaut-* [A *kot-*] 'to strike;'

\*wel- 'dominate' in Tocharian B *walo* [A *wāl*] 'king' from PIE \*w<sub>2</sub>l-ā-nt- or B *wlāw-* 'to behave oneself.'

The Tocharian word for 'seed,' AB *sārm*, is however not related to English *seed* or Latin *sēmen*. It is rather a deverbative noun to B *sāry-* 'to plant' just as *maim* 'measure' and *saim* 'refuge' are to *mai-* 'to measure' and *sai-* 'to support' respectively. In these three verbs we have a small series of deverbatives in \*-m̄n to \*-ye/o- presents whose root shows an ō-grade, i.e., *sōr-ye-m̄n*, *mōd-ye-m̄n*, and \**sōd-ye-m̄n*. Thus *sāry-* is to be analyzed as *sār-y-* and not *sā-r-y-* or the like.

3. Whether 'traditional' or 'statistical' all of these investigations have been limited for our purposes by their having been applied to Tocharian 'after the fact'—using data that may or may not be particularly pertinent to the placement of Tocharian.<sup>2</sup> The rest of this contribution will look at the problem of the connections Tocharian may or may not have with other Indo-European groups from a distinctly Tocharian point of view. To further this investigation we will discuss some new data, both 'traditional' and 'statistical,' dealing with phonology, morphology, and lexicon. As we will see the new data will sharpen our picture of Tocharian's relationship with other well-attested Indo-European groups.

### 3.1. THE FATE OF THE PIE SYLLABIC RESONANTS (\*R)

It is difficult on casual inspection to decide whether PIE \*R gave Proto-Tocharian (PTch) \*äR directly or whether the \*ä here is the reflex of one of the several short vowels that regularly gave late PTch \*ä, e.g. PIE \*u, \*a, or even \*i or \*e, though there is no trace of the palatalization of the preceding consonant which should have been produced if the vowel had been one of the latter. The origin of the \*ä is of some importance since the various Indo-European groups are divided by the "color" of the vowel which is generated by the PIE syllabic resonants: \*uR in Germanic, \*i/uR in Baltic and Slavic, \*uR (but \*eN) in Italic, \*Ri in Albanian, \*Ri in Celtic, and \*aR in Greek, Armenian, and Indo-Iranian.

In Proto-Tocharian an early \*aR or \*uR would have fallen together as \*äR everywhere except under two sets of circumstances: (1) after a word-initial \*k where \*u sometimes (but not always) shows up as late PTch \*wä rather than ä or (2) when immediately preceding another vowel—in such cases an early PTch \*aR would give later \*aR (cf. \*krAsniyom \*krasniyän → \*krañiyä(n) → krañi 'nape of the] neck' where PTch a from PIE \*A also came to stand in an open syllable). In this second kind of case, the reflex of a PIE \*R could appear directly before a vowel in early Proto-Tocharian only in the case where an intervening laryngeal had been lost, i.e. \*RHV → \*aRV or \*uRV.

In both cases where an early PTch \*a and \*u would give different reflexes, the evidence strongly suggests that the vowel was \*u. Thus where a PTch \*k precedes we have: AB kul- 'to fade away (of energy, etc.)' from PIE \*g<sup>(w)</sup>!H-o- [: Sanskrit glāyati, glāti 'he is languid, fades away'], A kuli [B kliye] 'woman' from PIE \*k<sup>(w)</sup>!HiEen- [: Mod. Irish caile 'country woman, maiden'], A kulmänts- 'reed; from PIE \*k!Hm- [: Greek kálamos 'id.'], AB kulyp- 'to desire, wish' from PIE \*k<sup>(w)</sup>!p-ye/o- [Avestan (aibi-)xrapa- 'to form a desire (for), etc.'], A kunmäsk- [B künmäsk-] 'to come; from PIE \*g<sup>(w)</sup>m-nA-ske/o-, AB kur- 'to grow old, fragile' from PIE \*grH-o- [: Greek géras 'gift of honor' from \*'privilege of age'], AB kursär 'mile' from PIE \*kr̥sru- [: Latin currus 'wagon'].

One can match this list with one of more or less equal size where neither Tocharian A or B shows any trace of \*w, e.g., AB kälp- 'to obtain' from PIE \*k<sup>(w)</sup>!p-, AB kän- 'to happen' from PIE \*g̑ñH-e/o, etc. Most of these are verbs where one may readily suppose that an original \*w disappeared by analogy with other forms of the verb where different ablaut grades existed, e.g., AB käl- 'to endure' [: Tocharian B kel- in the preterite], but a few morphologically isolated words also exist, e.g., A kánt [B kante] 'hundred' from PIE \*k̑ntóm. The presence or absence of a PTch \*w after an initial \*k is not predictable, either when comparing Tocharian as a whole with other Indo-European groups or when comparing Tocharian A with Tocharian B (note that Tocharian B never shows a \*w when A does not but that the reverse is often true). Nevertheless, in these cases a PTch \*u rather than \*a after an initial \*k seems sure.

As for the second case where an early PTch \*a and \*u would not yield the same reflex in later Tocharian, namely in combination with a resonant before a vowel (\*a/uRV), there is a nice morphologically isolated example in Tocharian B sprāne 'the two sides' from PIE \*sprHōn- [: Old English spura 'spur']. If the early PTch had been \*\*sparān-, the Tocharian B outcome would have been \*\*sparāne rather than the actual sprāne which presupposes an early PTch \*spurān-.

The Proto-Tocharian treatment of PIE syllabic resonants was then exactly the same as that found in Proto-Germanic: all \*R's become \*uR. This isogloss uniting Tocharian and Germanic is a highly significant one since in all likelihood it represents a shared innovation (partly shared also with Italic for \*r and \*l) and is thus

Van Windekens' (1976: 418) connection of sār-y- with Latin serere 'to sow' (from PIE \*'to sow in a row') is attractive, though the traditional derivation of the Latin word from a PIE \*si-sH-e/o- remains more likely.

<sup>2</sup> In large part because of the poorly understood nature of the Tocharian data. Of all of those offering opinions on the

relative placement of Tocharian, only Pedersen and Lane could be described as Tocharianists.

diagnostic for ascertaining the closeness of contact of the pre-Tocharians and pre-Germanics during some phase of late Proto-Indo-European.<sup>3</sup>

The treatment of the long syllabic resonants in Germanic and Tocharian is also identical. Both groups show complete, uncompensated, loss of the laryngeal in most cases. For Tocharian one should compare Tch *A kulmänts-* 'reed' from PIE *\*k<sup>l</sup>lHm-* with AB *kulyp-* 'to wish, desire' from PIE *\*k<sup>(w)</sup>lp-ye/o-*. Balto-Slavic shows *i/uR* for PIE *\*R* but *i/ūR* for PIE *\*R̃*. Italo-Celtic, on the other hand, has *Rā* normally, or *aR(i)* less usually, while Greek has *Rā/Rē/Rō*, depending on the color of the original laryngeal, or *aRa*. The identity of the Tocharian and Germanic treatments of the long syllabic resonants is striking but probably less diagnostic than the identity of the short syllabic resonants simply because the uncompensated loss of the laryngeal in this position could plausibly be an independent development.

### 3.2. THE GENITIVE DUAL

The collective evidence of Indo-Iranian and Balto-Slavic (albeit in the latter case only residually) assures the reconstruction of a PIE *\*-ows* in the genitive dual and a *\*-ow* in the locative dual. (Avestan shows an innovative genitive dual [as *if*] from PIE *\*-ōs*.) In addition, Indo-Iranian shows evidence of a genitive *\*-oyows* and locative *\*-oyow*, originally no doubt restricted to pronominal use. Greek also shows traces of a genitive in either *-ois* or *-oiois* from *\*(oy)ows* with either distant assimilation or conflation with the competing ending *-oi(i/u)n* (for forms and attestations, see Schwyzler, 1953: 557). While the ending itself has disappeared in Old Irish, the phonological remains of it in such cases as *fer* 'of the two men' with neutral final consonant and "aspiration" of the following word are compatible with a PIE reconstruction *\*-ows*, thus assuring the pan-Indo-European distribution of the genitive-locative *\*-ow(s)* and refuting any possible inference that it is an "easternism."<sup>4</sup> Its antiquity is also suggested by its general opacity—it is not formed on any obvious analogical model.<sup>5</sup>

<sup>3</sup> Thracian *may* show the same development of the short syllabic resonants (cf. Katičić, 1976: 145).

<sup>4</sup> No nominal duals are attested in Italic, Germanic, Anatolian, Armenian, or Albanian.

<sup>5</sup> One might speculate that the *\*-ow* is the endingless locative of the same stem one sees in the animate nominative-accusative, *\*-ōw* (with nominative lengthening as in *\*pAtēr*). The genitive is the same with the rather mechanical addition

Against this rather monochromatic background, the PIE *\*-oyisi/u-* seen in Greek *-oi(i)n/-oiun* (with permanent *n*-epithelystikon) and Tocharian B *-ais(-äñ/-i)*<sup>6</sup> stands out as a startling common innovation (Winter, 1962: 126). The Tocharian form seems to represent a PIE *\*-oyisi* since *\*-oyisu* would probably have given Tocharian B *\*\*oyis-* (cf. *soy* 'son' from *\*soyus* from PIE *\*suyus*) but in any case it would appear to be an old locative (with the plural ending *-si/-su*) added to the *\*-oy* that forms the neuter nominative-accusative singular. Such a rebuilding is, so far as I know, unique and thus highly probative of some special connection between Tocharian and Greek (pace Winter, 1962: 126).

### 3.3. "STRONG" AND "WEAK" ADJECTIVES

Perhaps the most striking innovation Tocharian shares with another branch of Indo-European is the possession of "strong" and "weak" (i.e., *n*-stem) adjectives at some time in its prehistory. Both Germanic and Tocharian evidently belonged to that portion of Indo-European, including at least Latin, Greek, and Balto-Slavic, where *n*-stem nouns became productive, as "singulatives" or definites, often with affective meaning, e.g., Greek *Strábōn* to *strabós* 'squinting' or Latin *Catō* to *catus* 'cunning, sharp' (see Maher, 1974: 148–49). In both Germanic and Tocharian *n*-stem nouns were apparently very popular and came to replace their non-*n*-stem counterparts, e.g., Gothic *augo* 'eye' or the word for 'tongue' in both groups, Gothic *tuggo* and Tocharian B *kantwo* (both reflecting PIE *\*dn̥ghwōn-*).

More important for our current purposes, in both pre-Germanic and pre-Tocharian, the singulative or definitizing function of the *n*-stems was extended also to adjectives—perhaps by way of the use of singulative nouns as nominal modifiers—creating parallel declensions of semantically definitive adjectives (morphologically *n*-stems) and semantically indefinite adjectives (those of other declensional types). Though the distinction between "strong" and "weak" adjectives had been effaced in the Tocharian languages by the time of their attestation, it had left numerous morphological traces in the declension of both nouns and adjectives (see also Adams, 1980). The presence of *n*-stem declensions of the comparative of the adjective in Baltic and Greek (compare the thorough-going "weak" nature of the

of *\*-s*. Note that these forms of the dual look like the corresponding cases of the singular rather than the plural.

<sup>6</sup> The Tocharian B *-äñ* and *-i* are superadded genitive endings (ultimately from a first singular pronoun *\*mene* and the *\*-eis* of the *i*-stems respectively.).

comparative in Germanic) may also be a part of this phenomenon, reflecting an early and prematurely curtailed stage of the extension of *n*-stem adjectives as definitive or singulative markers beside the non-definitive or non-singulative adjectives of other declensions.

### 3.4. THE MERGER OF HYSTEROKINETIC AND HOLOKINETIC *r*- AND *i*-STEMS

Not quite so diagnostic perhaps, since it might plausibly be thought to represent independent developments, is the merger of the hysterokinetic and holokinetic *r*- and *i*-stems in both Tocharian and Latin. For instance the paradigm of the word for 'father' (a hysterokinetic noun in Proto-Indo-European) shows the same hysterokinetic nominative paired with an originally holokinetic accusative in both Latin and Tocharian:<sup>7</sup>

	Latin	Tocharian B	Late PIE
Nom.	<i>pater</i>	<i>pācer</i>	* <i>pAtēr</i>
Acc.	<i>patrem</i>	<i>pātär</i>	* <i>pAtrm</i>

The same pattern is seen among certain PIE *i*-stems in both languages, e.g.:

	Latin	Tocharian B	Late PIE
Nom.	<i>plēbēs</i>	<i>āšce</i>	*-ē(y)s
Acc.	<i>plēbem</i>	<i>āśc</i>	*-im

Judging from the paradigm of the one extant *u*-stem adjective in Tocharian B, *colye* [nom. sg.<sup>8</sup>] 'wild,' *col\** [acc. sg.], *colyi\** [nom. pl.], *colām* [acc. pl.], Tocharian once had this pattern among *u*-stems also—an extension of this pattern that does not appear to have ever characterized Latin.

### 3.5. STATISTICAL COMPARISON OF LEXICAL RESEMBLANCE:

The number and kind of lexical resemblances joining two or more members of a particular linguistic family are of course important to determining the degree of interrelationship between the two branches. Various investigators of Tocharian have been quick to note individual isoglosses joining it with this or that other branch of Indo-European. So far such investigations

have been largely random and it is wise to check these intuitive choices by looking at whole groups of words so that the comparison might yield sounder results. For this purpose I have chosen two semantic fields dealing with "basic" vocabulary: (1) parts of the (human) body and (2) "biological terminology"—the designations for animals and plants and their parts. In addition, I took one semantically random group, namely those verbs which began with (Tocharian) *k*-. The three groups of words totaled some 365 words.

In each of these groups we are, of course, limited in our comparisons of those words which are attested in the surviving texts of the two Tocharian languages and to those which have a convincing Indo-European etymology but in all three cases there is a sufficiently large number of words to suggest a certain amount of confidence in the results. Resemblances in all three fields are "scored" the same way: complete morphological identity of the Tocharian word with its cognate in some other Indo-European branch gets a "3," root identity only scores a "1," and something that is in between in some way gets a "2." Thus Tocharian B *ānte* [A *ānt*] 'surface, forehead' is an exact match for Sanskrit *ānta-* which results in a "3" for Indic whereas it has only root cognates in Greek, Celtic, and Anatolian, giving each of those groups one point. Because unique resemblances, those shared by Tocharian and only one other Indo-European branch, would seem to be particularly important in determining degrees of relationship, I noted such cases with a plus and tallied the pluses separately (labeled "dyads" in the totals). Thus Tocharian B *pratsāko* [A *prātsak*] 'chest' matches Greek *prósōpon* exactly, except for the stem class, thus giving Greek a "2." Since no other Indo-European group shows the same compound (PIE \**prot-y-ōk*⁻), Greek gets a "2+." The totals for these comparisons are given in TABLE II.

TABLE II  
INDICES OF COMPARISON FOR LEXICAL RELATIONSHIP  
BETWEEN TOCHARIAN AND OTHER INDO-EUROPEAN GROUPS

	Ge	Ba	Sl	Ir	In	Ar	Gk	Al	La	Ce	An
Total points											
for three groups	96	65	48	34	63	2	73	9	35	34	11
of words											
Number of											
"Dyads"	16	8	4	0	4	0	8	0	2	2	2

<sup>7</sup> The evidence is ambiguous but the same pattern in the *r*-stems might be found in Germanic also.

<sup>8</sup> Found once, at K23 b3.

The rank order for closeness of overall lexical resemblance is: (1) Germanic, (2) Greek, (3) Baltic, (4) Indic, (5) Slavic, (6) Latin, Celtic, Iranian, (9) Albanian,

(10) Anatolian, (11) Armenian.<sup>9</sup> Independently of this investigation, van Windekens (1976: 614–619) has also tallied “concordances exclusives ou especiales” between Tocharian vocabulary and the vocabularies of other Indo-European groups. His ranked order is (1) Germanic, (2) Greek, (3) Indic, (4) Baltic and Iranian, (6) Latin, (7) Slavic, (8) Celtic, (9) Anatolian, (10) Armenian, and (11) Albanian. Excepting the last three, which in any case are to be grouped together in some fashion at the bottom of the list, the only serious difference between the two lists is the relative positions of Baltic and Iranian. The two lists reinforce one another to a remarkable degree.

4. No individual bit of phonological, morphological, or lexical data discussed here would be at all decisive as to Tocharian’s relationship with the other branches of

<sup>9</sup> One might suppose, however, that these results have been skewed by the differing rates of vocabulary loss that the various Indo-European groups have undergone. It is obvious, for instance, that both Armenian and Albanian have replaced a far higher proportion of their inherited word-stock than, say, Germanic or Greek. We can I think correct for such skewing, at least to a certain extent. Some thirty years ago Britton (reported in Ross [1950]) calculated the number of Proto-Indo-European roots, as recorded in Walde and Pokorny (1927–32), retained in some fashion in each Indo-European group. (Unfortunately he grouped Baltic together with Slavic, Celtic with Latin, and Indic with Iranian; there were no figures calculated for Anatolian or Tocharian.) The percentages of retention are: Celtic-Italic 63.7, Greek 62.6, Armenian 23.8, Indo-Iranian 54.6, Balto-Slavic 65.2, Germanic 67.5, and Albanian 15.6. From these figures we can calculate more nearly the original degree of lexical closeness of the non-Tocharian groups with Tocharian, the degree of relationship obtaining before the varying degrees of vocabulary loss obscured the picture. Thus for instance the total score (regular scores plus “dyads”) for Germanic is 112 which when divided by 67.5 (Germanic’s overall retention rate) is 165.9 which would have been the score presumably if Germanic had lost none of its Indo-European lexicon. The scores thus corrected are: (1) Germanic 165.9, (2) Greek 129.4, (3) Indic 122.7, (4) Baltic 112.0, (5) Slavic 79.8, (6) Iranian 62.3, (7) Latin 58.1, (8) Albanian 57.7, (9) Celtic 56.5, (10) Armenian 8.4. Albanian moves up a bit but otherwise the relative standings are much the same.

No matter what corrections we may make, we will probably never raise these figures to the level of statistical validity. The data doesn’t lend itself to that kind of precision. But the numbers are suggestive and, when coupled with other data leaning in the same direction, their suggestiveness becomes a presumption of correctness.

Indo-European. But when taken all together they point very strongly to a relatively close relationship that must have existed between pre-Germanic and pre-Tocharian dialects of late Proto-Indo-European. Next in importance are those resemblances which tie Tocharian with Greek; lesser ties existed between Tocharian on the one hand and Baltic, Latin, and Indic on the other. Such a position amidst other Indo-European groups is similar to that posited by earlier investigators though not quite the same—particularly the closeness of Tocharian with Germanic has not been emphasized before. Nor have investigators ever been struck by any particular resemblance with Indic, though it should be noted that this resemblance does not extend to Iranian.

It is possible, on the basis of these relationships with Germanic, Greek, etc., to “place” Tocharian geographically in the late Proto-Indo-European world in some manner, say, between Germanic (on the “north”?) and Greek (on the “south”?). It is also possible, and perhaps more realistic, to assume that the resemblances Tocharian shares with Greek, and perhaps also with Indic, are (when they are not simply common retentions) the result of later contacts in the immediately post Indo-European period of the pre-Tocharians with the pre-Greeks (and pre-Indics). These contacts would have occurred as the pre-Tocharians moved from a northwestern location near the pre-Germanics (and in the neighborhood of the pre-Latins and pre-Balts) to an increasingly southern and, ultimately, eastern position vis-a-vis the rest of Indo-European.

The supposition that Tocharian contact with Greek is later than the contact with Germanic receives some support from the fact that the pre-Greek with which pre-Tocharian speakers came into contact with seems already to have some specifically Greek features. For instance Tocharian B *pratsāko* ‘chest’ can be matched, except for the difference in declensional class, sound by sound with Greek *prósōpon* ‘face’. Both appear to reflect a PIE *\*proti-ōk<sup>w</sup>-o-* but the actual PIE form would appear to have been *\*proti-Ok<sup>w</sup>-o-* with a zero grade (cf. Skt. *prātīka-* or, with a different prefix, Latin *antīquus*). Only in Greek [: Greek *enōpē* ‘face, countenance’ *metōpon* ‘forehead’],<sup>10</sup> and in this one Tocharian form, do we see a generalization of the lengthened grade in this formation. The obvious inference then is that pre-Tocharian speakers borrowed the already transformed word as a whole from pre-Greek speakers. (It is also perhaps pertinent to note that only in this word has the Indo-European prefix *\*proti-* left any trace in Tocharian.)

<sup>10</sup> Cf. Hamp (1973:84).

It appears likely that the Greek-Tocharian contact occurred at a time when both groups were in contact with at least one non-Indo-European language. In this context we should note Winter's (1970) comparison of Tocharian A *nātāk* 'lord', *nāši* 'lady' with Greek *wánaks* (stem *wánakt-* from *\*wánatk-*) 'king' and *wánassa* 'queen' (from *\*wánakya-*). The un-Indo-European alternation of *\*-k-* and *\*-tk-* suggests an external source for these words and, given the meaning of the words, a source of some considerable prestige.<sup>11</sup>

To place the Tocharian contacts with Germanic, Greek, and other late Proto-Indo-European groups in time and space is of course very speculative. But it might not be too out of line to suggest that that variety of late Proto-Indo-European that was to become Proto-Tocharian first took shape and became in some measure distinguishable from other varieties of Proto-Indo-European in the course of the last half of the fourth millennium B.C. amidst the other members of Meillet's Northwestern Group (to which we should also add Albanian, whose closest ties are to Balto-Slavic [Huld, 1979: 258, and Hamp, personal communication]).

Geographically the Northwestern Group is surely to be located north of the Carpathians, somewhere between the Elbe and the Dniester. It is tempting to equate this Northwestern Group, defined linguistically, with the culture known to archeologists as the Globular Amphora culture which covered East Germany, Poland, Galicia, and Volhynia during the last half of the fourth millennium (Gimbutas, 1977: 303; see map, Gimbutas, 1980: 293).<sup>12</sup>

Within the Northwestern Group, Tocharian was closest to Germanic of the surviving Indo-European

groups, with secondary ties with Baltic and Latin. The Tocharians,<sup>13</sup> however, separated themselves very early from the rest of this group. They moved south and/or east and came into contact with another group of Proto-Indo-European speakers, the Greeks, perhaps in Moldavia or thereabouts, perhaps in the first half of the third millennium—certainly before the Greeks entered the Balkans proper. Subsequently the Tocharians must have continued to drift eastward across the north Pontic steppes and then Central Asian steppes, perhaps in this latter location briefly associating with some group of pre-Indic speakers,<sup>14</sup> ultimately to appear in history two thousand years later in Chinese Turkestan.<sup>15</sup>

In the absence of archeological data, the exact route of the Tocharians to their final home must remain somewhat speculative. The speculation is not however "idle," since it will be testable as our knowledge of Tocharian linguistics and Eurasian archeology becomes more complete. The "wanderings" of the Tocharians, whatever the exact details, are significant because they are important to our understanding of the interrelationships of the various branches of Indo-European and to our knowledge of the peopling of an at times significant, if always out of the way, corner of Asia.

Pontic steppes was curtailed. A group settling in a particular area found its contacts, linguistic and otherwise, restricted. Linguistic idiosyncracies were less likely to be leveled out and thus there was more divergence in speech. Conversely the steppe-dwellers lived in an environment where mobility was easy for a mounted people. Even though the steppes were large, constant interaction tended to efface incipient linguistic differences.

<sup>13</sup> Here as a shorthand for "speakers of pre-Tocharian" likewise for "Greek," etc.

<sup>14</sup> One should compare Thomas' supposition (1982: 81) that the pre-Tocharians may have formed one group of the Andronovo cultural complex, a complex which also included the Indo-Iranians.

<sup>15</sup> Gimbutas (1963) in her review of Bosch-Gimpera (cf. footnote 12) rejects his placement of the Tocharians in Volhynia because there is no linguistic or archeological evidence of a west to east movement across the Pontic steppes in either the bronze or the iron ages. I think I have provided here serious linguistic evidence for the western origin of the Tocharians. I don't think the absence of an archeological record is terribly damaging. Could we trace archeologically the historic migrations of the Slavs in Greece or, before them, the Vandals or Visigoths? In any case the movement of the Tocharians need not have been a mass migration, at least at first, but that of a relatively small group of the warrior elite that grew in strength only as they got east of the Volga, "Tocharianizing" their recruits in the process.

<sup>11</sup> It might be from this same non-Indo-European source that come Tocharian B *riye* 'city' and Thracian *bria* 'id.'

<sup>12</sup> Bosch-Gimpera (1961) places the Tocharians aboriginally in Volhynia at 1,500 B.C. or so. The time is surely too late but the placement may not be so far off.

One should note that of the well-attested Indo-European groups, seven (Celtic, Italic, Germanic, Tocharian, Baltic, Slavic, and Albanian) belong to this enlarged Northwestern Group. Four (Greek, Armenian, Iranian, and Indic) belong together in another group, one we might denominate "South-eastern." Anatolian stands apart from both. It is striking that so many Indo-European groups derive from one small portion, and a "colonial" one at that, of the total Indo-European area (which stretched at least from the Elbe to the Volga, and probably beyond).

The reasons I think are largely ecological. As some of the steppe-dwelling Indo-Europeans moved into the forested zone north of the Carpathians, the easy nomadism of the

## BIBLIOGRAPHY

- Adams, Douglas Q. (1980). "Towards a History of PIE *n*-stems in Tocharian." *Journal of the American Oriental Society* 100:439–443.
- Benveniste, Émile (1936). "Tokharien et Indo-Européen." In *Festschrift für Hermann Hirt*, Vol. II, pp. 229–240. Heidelberg, Winter.
- Bosch-Gimpera, P. (1961). *Les Indo-Européens: Problèmes Archéologiques*. Paris, Payot.
- Charpentier, Jarl (1917). "Die ethnographische Stellung der Tocharer." *ZDMG* 71:347–388.
- Gimbutas, Marija (1963). "The Indo-Europeans: Archeological Problems." *American Anthropologist* 65:815–836.
- Gimbutas, Marija (1977). "The First Wave of Eurasian Steppe Pastoralists into Copper Age Europe." *Journal of Indo-European Studies* 5:277–338.
- Gimbutas, Marija (1980). "The Kurgan Wave #2 (c. 3400–3200 B.C.) into Europe and the Following Transformation of Culture." *Journal of Indo-European Studies* 8:273–316.
- Hamp, Eric P. (1973). "Formations indoeuropéennes a second élément \*-(H<sub>o</sub>)k-." *Bulletin de la Société Linguistique* 68:77–92.
- Hermann, E. (1922). Review of E. Sieg's *Tocharische Sprachreste*. *KZ* 50:296–314.
- Huld, Martin E. (1979). *An Etymological Glossary of Selected Albanian Items*. Ann Arbor, University Microfilms International.
- Katičić, Radoslav (1976). *Ancient Languages of the Balkans, Part I*. (Trends in Linguistics, State-of-the-Art Reports, 4.) The Hague, Mouton.
- Kroeber, A. L. (1960). "Statistics, Indo-European, and Taxonomy." *Language* 36:1–21.
- Kroeber, A. L., and C. D. Chrétien (1937). "Quantitative Classification of Indo-European Languages." *Language* 13:83–103.
- Lane, George Sherman (1970). "Tocharian: Indo-European and Non-Indo-European Relationships." In *Indo-European and Indo-Europeans*, Cardona, Hoenigswald, and Senn (edd.), pp. 73–88. Philadelphia, University of Pennsylvania Press.
- Lévi, Sylvain (1913). "Le 'Tokharien B,' Langue de Koutcha." *Journal Asiatique* 1913.2:311–380.
- Maher, John P. (1974). "The Etymology of Common Slavic \*slověne 'Slavs.'" *Journal of Indo-European Studies* 2:143–156.
- Meillet, André (1914). "Le Tokharien." *Indogermanischer Jahrbuch* 1:15ff.
- Meillet, André (1967). *The Indo-European Dialects*. (Translated by Samuel N. Rosenberg.) University, Alabama, University of Alabama Press.
- Pedersen, Holger (1913). *Vergleichende Grammatik der Keltischen Sprachen*, Vol. II. Göttingen, Vandenhoeck und Ruprecht.
- Pedersen, Holger (1925). *Le Groupement des dialectes indo-européens*. Det Kgl. Danske Videnskabernes Selskab. Historisk-filologiske Meddelelser, XI.3.
- Pedersen, Holger (1938). *Hittitisch und die anderen indo-europäischen Sprachen*. Det Kgl. Danske Videnskabernes Selskab. Historisk-filologisk Meddelelser, XXV.1.
- Pedersen, Holger (1941). *Tocharisch*. Det Kgl. Danske Videnskabernes Selskab. Historisk-filologisk Meddelelser XXVIII.1.
- Pokorny, Julius (1923). "Die Stellung des Tocharischen im Kreise der indog. Sprachen." *Berichte des Forschungsinstitut für Osten und Orient* III:24–53.
- Porzig, Walter (1954). *Die Gliederung des Indogermanischen Sprachgebiets*. Heidelberg, Winter.
- Ross, Alan S. C. (1950). "Philological Probability Problems." *Journal of the Royal Statistical Society. Series B*, 12:19–59.
- Schwyzler, Eduard (1953). *Griechische Grammatik*. Fourth ed. Munich, C. H. Beck.
- Thomas, Homer L. (1982). "Archeological Evidence for the Migrations of the Indo-Europeans." In *The Indo-Europeans in the Fourth and Third Millennia*, Edgar C. Polome (ed.), pp. 61–86. Ann Arbor, Karoma.
- van Windekens, A. J. (1976). *Le Tokharien confronté avec les autres langues indo-européennes, Vol. I: La phonétique et le vocabulaire*. Louvain, Centre International de Dialectologie Générale.
- Walde, Alois, and Julius Pokorny (1927–1932). *Vergleichendes Wörterbuch der Indogermanischen Sprachen*. Berlin and Leipzig, Walter de Gruyter.
- Winter, Werner (1962). "Nominal and Pronominal Dual in Tocharian." *Language* 38:111–134.
- Winter, Werner (1970). "Some Widespread Indo-European Titles." In *Indo-European and Indo-Europeans*, Cardona, Hoenigswald, and Senn (edd.), pp. 49–54. Philadelphia, University of Pennsylvania Press.