Proceedings of the Sixteenth Annual UCLA Indo-European Conference

Los Angeles
November 5-6, 2004

Edited by:

Karlene Jones-Bley
Martin E. Huld
Angela Della Volpe
Miriam Robbins Dexter

Journal of Indo-European Monograph Series, No. 50
Institute for the Study of Man
Washington, DC
2005
Proto-Tocharian, Common Tocharian, and Tocharian — on the value of linguistic connections in a reconstructed language

Gerd Carling
Lund University

In this paper, some linguistic evidence for the positioning of Pre-, Proto- and Common Tocharian in time and space will be taken into account. The approaches have basically been selected in order to argue that the two Tocharian languages, Tocharian A and B, split up at a relatively late date, probably as late as the 1st century AD, after a long period of continuous linguistic change apart from the Indo-European language continuum. Finally, consequences and problems connected with these assumptions will be discussed.

0.0 The Textual Testimonies—a brief reminder

Tocharian A and B are attested from manuscripts and wall paintings from four ancient cultural areas along the northern fringe of the Täklîmanak Desert: the Maralbashi, the Kucha–Qyzyl, the Yanqi–Agni, and the Turfan–Chotscho areas. Tocharian B, or “West Tocharian,” documents are found at all sites, in total around 3,200 documents, many of them very fragmentary. Tocharian A, or “East Tocharian,” is found only at sites in the east (Yanqi, Shortuq, Chotscho, Toyuq, Murtuq, Sängim, and Bázâqliq), in all around 500 documents. Very scanty are the remnants of a third Tocharian language, Tocharian C, found in Niya Praktik documents from sites along the southern fringe (Loulan, Niya, etc.) from the 3rd century AD.

Some of the Tocharian B caravan passes can be dated with great certainty to the first half of the 7th century AD. The oldest Tocharian manuscripts, found in the Kucha area and written with an archaic ductus, might be traced back to the 6th or even the 5th century AD (Pinault

---

1 I thank Ken-ichi Takahama, Victor H. Mair, Georges-Jean Pinault, Wolfgang Behr, Seishi Karashima, and the SCASSS seminar for valuable remarks. I thank Judith Josephson for correcting my English.
The meaning ‘old’, ‘geriatric’ is an innovation of the non-Tocharian languages (Carling 2003b:92).

In a recent project, conducted by Don Ringe, Tandy Warnow, Ann Taylor, and others (see Ringe et al. 1998, 2002), cladistic computer modelling is used to subgroup linguistic features of the Indo-European languages. Input data consists of 22 phonological items, 15 morphological items, and 333 lexical items (based on the Swadesh-word list of Tischler 1973). The project, being a more detailed and advanced version of the glottochronological method, especially with a more adequate selection of input data (cf. the Appendix of Ringe et al. 2002), is interesting in so far as it manages to reproduce a relatively stable evolutionary tree for Indo-European. Of importance for our purpose is that the position of Tocharian, as being one of the earliest branches to split off from the Proto-Indo-European tree, seems to be confirmed even by this independent method (see Ringe et al. 2002, fig. 8; Gray and Atkinson 2003).

1.2 On the value of typological innovations

Despite the archaisms mentioned in the previous section, Tocharian has a structure which is different from many other Indo-European languages. The transition from inflection to agglutination and group inflection has been regarded as a non-Indo-European influence (cf. Thomas 1985:61-62, and especially the works by K.H. Schmidt 1982, 1987, 1988). Without going into great detail, I will give an overview of the most prominent typological innovation of Tocharian: the reorganization of the case paradigm.

The reorganization of the case paradigm is relatively easy to detect, even though the finer points, especially on the semantic side, have not been solved completely (see Carling 2000:381-388). The Indo-European case system collapsed in Proto-Tocharian, most likely at an early date and due to the loss of final syllables. However, the loss of final syllables (at least the dropping of final vowels) must have been operative for quite a long period, since we know that this sound change affects, for instance, the word for Buddha > Cm.Toch. *put > *pdt (see 3.1), as well as other contemporary or later Indo-Aryan borrowings independently in Tocharian A and B, up to a certain period (see 3.2).

The remnants left behind after the collapse of the case system constituted a rudimentary system of three cases: nominative, oblique, and
a genitive-dative. In this arrangement, the ancient oblique was also partly lost and replaced by a new marked category: -η, used for the oblique of sentient beings. Quite likely, case functions at this early stage were expressed by using postpositions. At the next level, a case paradigm with two layers emerged, constituted by one layer of inflectional elements, the so-called "primary cases" (nominative, oblique, and genitive) and one layer of affixes which used the oblique as a basis for their attachment, the so-called "secondary" cases. Three secondary cases can be reconstructed for Common Tocharian: perative *-t, allative *-cā, and locative *-nē. In Tocharian A, the thematic vowel *x was unverbated with the affix, giving the forms perl. -x, all. -xc, and loc. -xām. In Tocharian B, the *x of the oblique plural ending *-ns was unverbated with the affix, giving the forms perl. -s, all. -sc, and loc. -nē (cf. Klingenschmitt 1975; Gippert 1987). Somewhere at this level, the genitive might have been vacillating between the groups, which parts of the morphology of the genitive indicate (for an overview of the genitive morphology see Pinault 1989:87-89, 98-99). After the split of Tocharian A and B, new cases were added to the paradigm, an ablative Toch.A -dš, a comitative Toch.A -aššiā B -mpa, an instrumental Toch.A -yo, and a causal Toch.B -ni. This means that we can reconstruct at least four stages preceding the case paradigms of Tocharian A and B (cf. Carling 1999:96): 1) A rudimentary Proto-Tocharian paradigm with only three cases (nominative, oblique, and genitive) covering the most basic grammatical functions, 2) a Common Tocharian agglutinative paradigm consisting of three inflectional and three agglutinative cases, 3) a period of confusion concerning stem and affix, resulting in a univerbation, probably coincidental with the split of Tocharian A and B, 4) two separate Pre-A and Pre-B paradigms extended with new, agglutinative affixes.

The case system in Indo-Aryan had a parallel evolution, which also partly extended over historical periods and is therefore of interest for our purposes. Following the description by Masica (1991:230-248), the inflectional system of Old Indo-Aryan, i.e., Vedic and Sanskrit, which was very similar to the system reconstructed for Indo-European, began to break down in Middle Indo-Aryan, i.e., Prākrit and Pāli. The breakdown, "a result both of internal pressures and of phonological erosion" (Masica 1991:231), resulted in a very simplified system in Late Middle Indo-Aryan, i.e., Apabhraṃśa. Thus, the system developed in Indo-Aryan was very similar to that of Tocharian, even if the systems in the separate dialects later gradually became more complicated. In the reconstruction of Masica the case system is stratified into three layers. Layer I corresponds to the remnants of the Old Indo-Aryan inflectional system. Layer II is built by affixes or particles that are attached either to the base directly or to a Layer I element. Layer III is by definition mediated by a Layer II element, most normally a genitive. However, in many New Indo-Aryan languages, this third layer is optional (for an overview see Masica 1991, figs. 8.6-8.8). Romany, a New Indo-Aryan dialect that most likely separated from the Indo-Aryan linguistic continuum at a very early date, shows a relatively uncomplicated system, which has striking typological similarities with Tocharian: Layer I consists of a nominative and an oblique, and Layer II is constituted by affixes attached to the oblique stem (see e.g., Boretzky 1993:30ff).

What conclusions can be drawn from these systematic restructurizations being parallel in Proto-Tocharian and in Indo-Aryan? I think there is reason to suspect that a parallel investigation of the two paradigms, from a morphological as well as from a semantic perspective, would give interesting results. However, of importance for our purpose is the chronological perspective: in Indo-Aryan, the decay and restructuring of the system was extended over a period of more than a thousand years. The decay of the Old Indo-Aryan paradigm, triggered by similar processes as the Proto-Tocharian decay, did not take place in a moment; the fact that it was extended over a quite long period is something that we also might suspect for Proto-Tocharian. In Indo-Aryan, the restructuring began with the formation of the New Indo-Aryan dialects and continued separately after the separation of the dialects and dialect groups; this we notice for Tocharian also: A and B developed and completed their paradigms independently, but in a similar direction.

Furthermore, there seems also to be reason to suspect that the verbal system of Tocharian was restructured, not in a similar way, but to a similar extent as was the case paradigm. The verbal system, which is extremely complicated and distinguishes 12 different classes of the present, 11 of the subjunctive, and 6 of the preterit, is composed of Indo-European elements and categories. However, synchronically regarded, the selection and classification of different categories seems to be determined by factors such as valence and aktionsart, which is of course not a surprise, but which on the whole seems to be an innovation proper to Tocharian. The details have not been investigated completely (see Carling in press).
If we consider large-scale typological innovations such as those mentioned above, it might be tempting to try to reconstruct or refer to areal linguistic contacts in prehistory. Nonetheless, Tocharian and Indo-Aryan had, as we have seen before, almost parallel evolutions, which took place under completely different circumstances and during different periods. Therefore, typological changes should be avoided for the reconstruction of language contact in prehistory—the safest indicator of language contact in prehistory is, by all means, good reconstructions from vocabulary and phraseology, as will be discussed in the next section.

2.0 Language Contact in Proto- and Common Tocharian

2.1 Traces of early language contact

Language contact preceding the state of Common Tocharian cannot always be reconstructed with great certainty, especially because of the large-scale phonetic and phonological changes and reductions in Proto-Tocharian, above all the devoicing and de-aspiration of the stop series. Relatively numerous proposals of borrowings from various sources into Tocharian as well as borrowings from Proto-Tocharian into other languages should be regarded with some scepticism—especially if language contact is not historically expected. However, judging between semantic and phonetic matching (cf. Lubotsky 1998:381), I think that a perfect semantic matching is of greater importance, especially if one considers, on the one hand, the large-scale phonetic and phonological changes in Tocharian, and, on the other hand, the problems existing within the reconstruction of pre-forms of other languages, e.g., Chinese.

Concerning very early contact between Tocharian and neighboring languages, it is worth noticing that Tocharian does not seem to have received any material from Common Indo-Iranian. This kind of “negative evidence” is not, of course, of the same value as real evidence of language contact, which we have from both Indo-Aryan and Iranian. On the other hand, this fact says something either about the coming of the Tocharians into Central Asia, or the coming of the Indo-Iranians.

However, there seems to be a number of very early borrowings, probably dating back to the 2nd millennium BC, which co-occur in Indo-Iranian/Early Indo-Aryan and in Tocharian. The donor language is neither Indo-Aryan nor Tocharian, since all these examples lack convincing etymological explanations in both languages. Presumably,

the donor language was some (at least for the moment) unknown Central Asian language.

• Toch.B *aŋi ‘hip, groin’ translates Skt. śrōṇī ‘the hips and loins, buttocks’ (MW:1102c) in a bilingual text (Carling 2003a:48), which is compared by Pinault (2003) to Skt. aṇī ‘linch-pin, peg, bolt’ or ‘the part of the leg just above the knee’. Pinault assumes that both words are borrowed from an *aṇī meaning ‘hip’ → PToch. *aṇi. Since aṇī is already attested in Vedic Sanskrit, supported by the fact that the sound change PToch. *a > Toch.B a [A a] does not apply to later Indo-Aryan borrowings (cf. below), we must assume a very early acquisition.


• Toch. B ʃecca, A ʃišak ‘lion’ is an old and much-debated question. The connection with Skt. śīna- (also śīṇaka-) ‘lion’ (MW:1213) has been interpreted as both Indo-European heritage (including also Arm. inž ‘leopard’) and an early borrowing from Indo-Aryan to Tocharian, or an acquisition from an unknown language to Indo-Aryan and Tocharian (for a summary see Behr 2004:32-34). The word occurs also in Chinese,2 in two different forms, Mod.Ch. suānǐ, Mid.Ch. *suān-i, OCh. *suɔf+j+i+nee (GSR 46d+8730, 8730#) Tib. se-n-ge ‘lion’ vs. Mod.Ch. stīzī, Mid.Ch. *stij+r+i+to (GSR 559a, 559#, 964a) ‘lion’. These words cannot be connected etymologically; therefore, they most likely represent different stages of borrowings (see Behr 2004). Since the words on all

---

2 A note on Chinese reconstructions: Modern Chinese (Mod.Ch.) refers to Pīnyīn, the Old Chinese (OCh.) reconstructions to the system used by Baxter (1992), including the revisions by Sagart (1999). The space after a GSR number indicates that the character does not exist in GSR, but has a phonetic value which can be assigned to this GSR series. Deviations from titles are noted. I thank Wolfgang Behr and Ken-ichi Takashima for helping me with these transcriptions and reconstructions.
sides seem to lack good etymologies, the idea of an unknown donor language is quite appealing. Furthermore, the suffixation of Cm.Toch. *-kar might be compared to the previous example, Toch.B *šcen / *šcak A *išāk (see also Pinault 2002:330). Considering the problematic phonology of all the variants, it is likely that we are dealing with two different (but probably related) words, designating (different species of?) large felines. The Tocharian forms were most likely borrowed independently, after the split of Tocharian A and B (see Behr 2003:35).

* Toch.B kercapo ‘donkey, ass’ is regarded as a very early borrowing from early Indo-Iranian, Skt. garabha- ‘donkey, ass’ < *gord(h)ebho-, taking place before the merger of PIE *a, *e, *o > *a in Indo-Iranian (see Adams 1998:195-196 with further references). Being for the moment the only example of this borrowing and supported by the fact that Skt. garabha- lacks a good Indo-European etymology, I think that it is reasonable to include also this word among the borrowings in Early Indo-Iranian and Proto-Tocharian from an unknown Central Asian language.

Another topic for discussion is the presumed presence of early Turkic borrowings in Proto-Tocharian, as presented by Lubotsky and Starostin (2003). The idea came up with the connection of Toch.B kaum, A kom ‘sun, day’ with Uygh. kün, Turkm. gün etc. ‘sun, day’ by Pedersen (1944:11) and Winter (1963), interpreted by Pedersen as a borrowing from Turkic into Tocharian, by Winter, interpreted as the opposite. Adams excludes any connection with Turkic on historical grounds (Adams 1998:211): “To have given both A kom and B kaun, the borrowing would have had to have been of PToch. in date. So early a date might itself rule out the ‘Turks on geographical grounds.’ However, if the terminus of Common Tocharian is situated at a later date, as in this paper (cf. also Pinault 2002:262), Turkic borrowings would seem less improbable. Even if Lubotsky and Starostin’s list of Turkic borrowings contains several problems—as Toch.A āle, B alyive* ‘palm (of the hand)’—of which the Indo-European etymology seems to be the most convincing proposal (see Adams 1998:27), or Toch.A tmäm, B iqmämne ‘ten thousand, a myriad’, which seems to be a Central Asian migratory term, occurring in Altic languages as well as Iranian and (probably also)

Chinese, and of which the origin remains very uncertain (cf. Van Windekens 1976:642)—other etymologies cannot be disregarded:


As demonstrated by Pinault (1998:359-360), the Turks took over, probably from early Tocharian A, the pre-Buddhist concept of ‘name and glory’, Toch.A nöm-klyu B nöm-kálywe < Cm.Toch. *nöm-klywe (cf. Ved. nāman- śrávas-, Hom. ḫwāmā klēfōc). Otherwise, Turkic words in Tocharian texts are restricted to names and titles in colophons (cf. Winter 1963), of which some also occur in Iranian languages. This would be a relatively late phenomenon.


Also late is the linguistic influence of Tocharian A on Old Uyghur, which was abundant, especially due to the translation of Tocharian A texts into Old Uyghur (cf. Winter 1963; Pinault 2001b).

Language contact between Tocharian and Chinese has attracted the interest of scholars for several years. Apparently, the exchange of loanwords went in both directions, from Proto-Tocharian into Old Chinese and from Old Chinese and Middle Chinese into Proto-Tocharian and Early Tocharian A and B.

- The most certain borrowing from Tocharian to Chinese is the word for ‘honey’, Toch.B mīt, mīt < *myit < *mī’t < IE *méd ‘honey, mead’ (Porkorny 1959:707), Mod.Ch. mi, Mid.Ch. *mjīt, OCh. *mjīt (GSR 405r) ‘honey’. The Middle Chinese reconstruction corresponds precisely to Cm.Toch. *myit.
Lubotsky (1998) lists a number of Tocharian loan-words into Old Chinese, some of which are worth considering, especially since they fall within the same range of meanings: terms for chariots and chariot gear.

However, except for his proposals for terms for “town building” remain, except for Ch. li ‘village, hamlet’ which will be discussed in further detail below, his proposals still remain uncertain since they, on the Tocharian side, refer to either relatively vague verbal roots, or the semantics are not close enough (cf. Ch. li ‘wall’ and Toch. B pkants, A pkants ‘obstacle’).

In a recent article, Lubotsky and Starostin (2003:264) divide Chinese borrowings into Tocharian in two distinct groups: 1) Early loans “showing pre-Han or Early Han peculiarities” and 2) Middle Chinese loans. The examples of borrowings from Old Chinese, four in number in their list, are of interest for our purposes and therefore worth a closer look.

- Toch. AB klu ‘rice’ was likely borrowed from Old Chinese: Mod. Ch. dào, Mid. Ch. *dauX, Och. *C-luu-ton ‘rice, rice-paddy’ (GSR 1078). In Middle Chinese, the initial cluster Och. *gl- was simplified to -d- (Baxter 1992:232).
- The derivation of Toch. B rapanše ‘of the last month of the year’ in LP3 12 a2 rapanše menne ikän-wine ‘on the second day of the month rapanše’ seems also persuasive. The adjective is most likely formed on a noun *ræp, which might be borrowed from Old Chinese: Mod. Ch. là, Mid. Ch. *lap, Och. *rap (GSR 637) ‘winter sacrifice’. It is not unlikely that an earlier meaning of the Chinese word is reflected in Tocharian.
- Toch. A ri, B riye < CM. Toch. *riye matches the Old Chinese reconstruction of Mod. Ch. lì, Mid. Ch. *liX, Och. *pr-7 (GSR 978a) ‘walled city’. However, it is a question whether this word was borrowed from Chinese into Tocharian or vice-versa. The contra is a possible Indo-European etymology *wriH-eH, occurring in Thracian βπας (see Lubotsky 1998:386, cf. also Mallory and Adams 1997:210). The Old Chinese word is attested already in Zhou I (11th–9th centuries BC), which means that this

---

3 LP denotes Laisger Passer de Caravannes, i.e., “Caravan Passes” and is the generally accepted abbreviation used for the caravan passes of the Paris collection. Full information is found in Lubotsky and Starostin (2003).

may have been a very early borrowing, if it was overtaken from Tocharian. The pro is the fact that the word for ‘(smaller) village’, Toch. A suks, is also a borrowing from Chinese (see below), and if the Tocharians borrowed one term for ‘village’ of a certain size, it is also likely that they borrowed another one.

- However, more questionable is the derivation of Toch. A tronik, B troik plkh ‘hollow, cave’, as originating from Mod. Ch. zong, Mid. Ch. trjwung, Och. *trug (GSR 1078) ‘middle’. The meaning in the Old Chinese literature is clearly ‘middle’ and ‘empty, hollow’ is found in Shuowen jiezi only, which is of uncertain value, since the interpretations in this early Chinese dictionary (100 AD) very often are influenced by the interpretation of the forms of the characters themselves. An alternative might be Mod. Ch. dòng, Mid. Ch. *dungh, Och. *looy-[s] (GSR 1176h, 1176l) ‘cave, grotto’, which then ought to be borrowed during a post-Old Chinese period.

In dating these sound changes from Old to Middle Chinese, it is commonly accepted that the simplification of initial clusters may be traced to around the time of the Early Han, i.e. last two centuries BC. For the dating of the transition from *r > *l, it is likely that this process did not happen synchronically in all dialects. Pan (2000:270–271) believes that this transition was fully completed by the late 4th century AD.

Later borrowings, from Early Middle Chinese, are relatively abundant and cover certain semantic fields: economy, techniques, and institutions. Some of them are shared with other neighboring languages, such as Sogdian and Khotanese. Good overviews (with references and discussion) are given by Thomas (1985:153–154) and Lubotsky and Starostin (2003:262–265). Some others can be added:

- Toch. A sostuŋk ‘tax collector, banker’ (Skt. śreṣṭhaṁ) corresponds to Niya sosthama ‘tax collector’, Bactr. sostaygo < *soṭanggo. The origin of the Tocharian word is uncertain, with either Niya Prakrit or Buddhist being likely sources. A possible source for all variants is a combination Mod. Ch. shōu-cāng, Mid. Ch. *suwNZx-dzāng, Och. *xiw-N-s-[h]raŋ (GSR 1103a+727g) ‘receive, accept, gather’ + ‘conceal, store’ (Finaut 2000, apud Nicholas Sims-Williams and Victor H. Mair, pers. comm.).
Toch.A șuk ‘(smaller) village’ (for examples see Poucha 1955:347), is paralleled by Toch.B kwao ‘village’. The attempts to connect these words and give them an Indo-European etymology (see Adams 1998:234-235 and Hillmarsen 1996:197-198) remain unsatisfactory. Another proposal for Toch.A șuk is Mod.Ch. su, Mid.Ch. *sjwok, OCh. *suk (GSR 1029a) ‘lodging, mansion’. Itô and Takashima (1996:1401) reconstruct OCh. *sjok* (with a final *s (that has a function of localization and production of nomina actionis, etc.), which became lost around the 3rd-4th centuries AD and replaced by tone.

Toch.A ąkk ‘seal, stamp’ in 265 al ąkk̯ąmsal piksit volkar ‘they have ordered to write [it] together with a seal’ is most likely borrowed from Mod.Ch. yín, Mid.Ch. *ʔjwInH, OCh. *ʔjIn-s (GSR 12511), *ʔjIn (Takashima) ‘seal, stamp’. Whether the final -k is a result of the Old Chinese palatal *ʔ or a Tocharian addition remains uncertain (the outcome of *ʔ might also be a velar *ʔ, which would yield Toch. *əK > Toch.AB ǝk).

2.2 The case of Indo-Iranian

Since the stock of Iranian borrowings is of importance for the dating of Tocharian, a brief overview of the different stages of Iranian borrowings will be given here (following the basic design as constructed by Krause 1955 and further developed by Winter 1971; K.T. Schmidt 1985; and Pinault 2002; cf. also Isebaert 1980).

The earliest contacts between Tocharians and Iranians date back before the beginning of our era and should be linked to a language which is, obviously, Old Iranian rather than Middle Iranian. Since our knowledge of Old Iranian is limited to Avestan and Old Persian, which do not give perfect matches, the donor language was probably some other unknown Old Iranian dialect; a language ancestral to Ossetic has been proposed (Pinault 2002:245). There are a handful of convincing etymologies, as collected by K.T. Schmidt (1985) and Pinault (2002).

- The Old Iranian diphthongs ai and au, which are monophthongized in Middle Iranian and in Old Persian already before Darius (see K.T. Schmidt 1985:766, n.10), are kept in Toch. B waipcee ‘possession’, from OIranian, Av. x’āparīya- ‘own’ (cf. also Winter 1971:218) and Toch.B waipite (with variants waiptār, waiptāyar) ‘separately, apart’ < Cm.Toch.

*waip-pāta, borrowed probably from an adjective, OIran. *xwai-pat in the sense of ‘independent, oneself’ (Pinault 2002:246f.), as given also in Av. x’āparīya- (cf. above). A further example is (K.T. Schmidt 1985:763) Toch.B lisan ‘arrow’ (Skt. lsAN) from an OIran. *dzaiina-, Av. zaēna- ‘weapon’. This form is interesting also from the perspective that the affricate *dz, in Tocharian devolved as *ts, indicates either that the borrowing took place at a very early date or that the donor language had conservative phonological traits.

The Middle Iranian borrowings can be divided chronologically into two groups: 1) Bactrian and 2) Sogdian and Khotanese. The Bactrian borrowings are obviously older than the other Middle Iranian borrowings.

The Bactrian borrowings require some reconstruction, since many of them are affected by phonetic change, which is different in Tocharian A and B. Moreover, the stage of the Bactrian donor language normally also has to be reconstructed, even if the Bactrian traits seem quite inevitable, as was demonstrated first by Winter (1971) and has been developed in further detail by Pinault (2002). The most prominent feature of the Bactrian borrowings into Tocharian is the retention of the Old Bactrian trisyllabic structure.


- Toch.A āpātrīk ‘craftsman’ (Couvreur 1955/56) is likely a borrowing from Bactr. āpātra ‘master (craftsman)’ < *api-sī+a- (Sims-Williams 2000:175). The proposal Ājīvika’ by Sieg (1952:34) is phonetically improbable. The phonetic evolution of this complicated word would then be reconstructed as Cm.Toch. *ap(a)ṣtat > *apṣt(a)ī + ik > Toch.A āpātrīk. The *t would likely become lost in the consonant cluster * pst. The insertion of
This means that the Bactrian borrowings entered Tocharian at a time when the Bactrian final vowels were not mute (which they were at the time of Kaniška), and when Tocharian A and B were still not separated. Most likely, Bactrian borrowings were adopted during a period when the Kushan empire was strong and influential and also expanded its interest into the Tarim Basin. Sometime around the beginning of our era is most likely the period of these adoptions, which also means that Tocharian A and B at this moment were still not separate (cf. Pinault 2002:262; K.T. Schmidt 1985:765).

There are indications that even Khotanese borrowings entered into a Common Tocharian state:

- An example is Toch.A *misi, B misa 'field' (not 'Gemeinde' as in Krause and Thomas (1964) < Cm.Toch. *miṣa, borrowed from Khot. mūṣa, miṣa- 'field' (see Pinault 2002:267).

3.0 The Post-Common Tocharian Period: Buddhism and Indo-Aryan influence

3.1 Rendering of Buddhist terminology in Tocharian A and B

The period following the split of Tocharian A and B, whenever this split took place (see the discussion in 4.0), is most predominately characterized by the acceptance of Buddhism and the huge influence of Indo-Aryan, basically Sanskrit. Considering the Buddhist vocabulary, it seems quite evident that the rendering of the terminology, i.e., translation and adaptation of the original Prākrit or Sanskrit terms, was done separately in Tocharian A and B. One of the most basic components of Buddhist faith, 'the Law' (Skt. dharma) Toch.A mārkampal, B pelaiyne, was formed differently in the two languages, by using the same element Toch.B pele, A pal 'law' < Cm.Toch. *pele (probably of Indo-European origin, see Adams 1998:398).

- Toch.B pelaiyne < *pele-yākne was constructed from Pre-B *pele plus Pre-B *yākne in Toch.B yakne, A wāṃ 'manner, way of doing', of Indo-European origin (see Adams 1998:481). Toch.A mārkampal probably constructed from the verb


Another very basic concept, the 'community', Skt. samgha-, was borrowed, probably independently, into Toch.AB sāṅkī.

However, the word for Buddha himself, Toch. A ptaṅkāt, B paṅkāte, has its source in Common Tocharian. Skt. buddha- was borrowed into Common Tocharian and adapted as *pāt < *puti and was thus affected by a sound change operating on Indo-European words, e.g., Toch.A rātur, B ratre 'red' < *rātūr < *(H)ruðhr-rö. This sound change continued to operate in Tocharian B and affected other borrowings from Indo-Aryan (cf. below), but ceased to operate in Tocharian A, which is demonstrated by the equivalents of Skt. sukt- 'happiness', Toch.A suk, B sakw, Cm.Toch. *pāt is present in the B word for 'Stūpa', pat. Thereupon, Tocharian A and B independently formed their word for 'Buddha' by using the same construction, typical for names of gods, indigenous as well as adopted ones (cf. Pinault 1998:358-359), i.e. *X-nākta 'the God X', Pre-A *pāt-āṅkta > *pāt-āṅkāt > Toch.A ptaṅkāt and Pre-B *pāt-āṅkte > Toch.B paṅkāte (cf. Pinault 1995:19-20, see also Kim 1999:115-120 on the chronology of the sound changes in the Tocharian word for Buddha).

Apart from this, the Buddhist vocabulary shows much of the same pattern as does the vocabulary described in the next section. We find occurrences of borrowings from Sanskrit beside literal translations. The vocabulary is predominantly borrowed from Indo-Aryan, but exceptionally also from Iranian. The Indo-Aryan borrowings represent all stages of sanskritization, from Middle Indic forms to pure Sanskrit (see Pinault 1995).

3.2. The process of sanskritization as reflected in Tocharian:
from Gāndhārī to Buddhist Sanskrit

In an attempt to stratify the Indo-Aryan material borrowed into Tocharian, we may, tentatively, distinguish at least three main stages, of which the Sanskrit group can be sub-stratified into more chronological stages: 1) a source in Gāndhārī, 2) a source in some local (Dardic?) Middle Indo-Aryan dialect, 3) a source in (Buddhist) Sanskrit or in Sanskrit a) with the effect of a number of sound changes characteristic of Tocharian A or B, among others devoicing and deaspiration of stops (in all positions), b) with the effect of devoicing of stops in final position
only and loss of final syllable in both Tocharian A and B, c) Sanskrit final -a > Toch.B -ae, A -Ø.

It is not self-evident that these stages should be seen as strictly chronological, neither for Tocharian nor for Indo-Aryan. We could expect a number of cross-occurrences or forms revealing features from several groups. On the whole, the stages also reflect the process of sanskritization of the Sanskrit language itself, from prakritized forms to relatively "pure" Sanskrit. This results in a pattern where later borrowings in Tocharian are much more "Sanskrit-like" than earlier ones, not only because they are affected by fewer phonological changes but also because their source is a more classical Sanskrit form.

Indo-Aryan loans are extremely abundant in the Tocharian literature; therefore, this brief survey of borrowings representing the different stages mentioned above only constitutes a small minority of the total amount of borrowings. Much of the vocabulary has a technical nature, i.e., reproducing terms within the spheres of religious worship, and legislation, medicine, astrology, and other disciplines. Here, the basis of the investigation has been a couple of medical texts (PK.AS.2A, 2B, 2C containing the Yogaśataka, bilingual Sanskrit–Tocharian B, re-edited by Carling 2003b), which have a highly technical vocabulary, including examples of terms for diseases, cures, medical ingredients, various plants, and parts of the human body, and therefore provide a large amount of borrowings from Indo-Aryan. The medical examples are then extended by other, typical examples of the group mentioned.

1) A source was Gandhāra, the administrative language of the Kroraina kingdom (3rd century AD) and most likely the official language of the first wave of Buddhist missionaries into the Tarim Basin (see Pinault 2000, 2002). Interestingly enough, the Gandhāra terms as borrowed into Tocharian are not very abundant; probably this "first wave" of Indo-Aryan vocabulary into Tocharian was not as overwhelming as the later one, when Buddhist Sanskrit terms flowed into the language.

- Toch.B puščə (root of the) Saussurea lappa' in PK.AS.2A b3 corresponds to puškaramila (puška- 'lotus; Costus speciosus' (MW:638e) and māla- 'root') of the Sanskrit bilingual text. The source of Toch.B puščə is obviously Pkt. pūsa-, Pāli pussa- 'flower' (from Skt. puṣya- 'vigor' Turner 1966:470b).

- Toch.B gurucī 'Tinospora cordifolia', borrowed from an equivalent of Skt. guḍucī 'Coccus cordifolius' (MW:357a), shows an early rhotacism of Skt. ḍ > r, ḍ, which occurs in New Indo-Aryan, but which was most likely already present in Gandhāra. Elsewhere, Middle Indo-Aryan has l, l (von Hinüber 2001, §201). Toch.B arirāk 'Terminalia chebula' from Skt. haritaki 'the yellow Myrobolan tree, Terminalia chebula' (MW:1292a) reflects the same type of transition.

- Toch.B ṭesph 'phlegm' (Skt. kapha-) most likely had a form of Skt. ślesman- 'mucus, phlegm' as its origin. The Middle Indo-Aryan forms (Pāli silesuma-, Pkt. silema-, sīlima- 'phlegm') look different. However, a transition -sm- > -ṣm- > -sp- would be possible in Gandhāra (cf. the New Indo-Aryan dialects: Kho. solasp, Kal.numb. sīləs < *sīləsp, Turner 1966:738b).

2) Another source may have been some local Middle Indo-Aryan dialect, perhaps Dardic. Although Gandhāra has a number of Dardic traits (see von Hinüber 2001, §93), some of the early Indo-Aryan loanwords in Tocharian cannot be explained as borrowed from Gandhāra exclusively. Therefore, it would be possible to assume that some other Indo-Aryan dialect, different from Gandhāra (and Sanskrit), conveyed vocabulary into Tocharian. This is where we can put examples of Indo-Aryan borrowings not fitting into the regular Gandhāra or Sanskrit pattern.

- A typical problematic example is Toch.B kāṣu 'Saussurea lappa', translating Skt. kuṣṭha- 'idem', which is obviously also the origin of Toch.B kāṣu. The Skt. ḍṛ(h)-cluster typically results in Middle Indo-Aryan ḍṛ(h) (von Hinüber 2001, §229), Pkt. kuṭha- 'a particular drug', Pāli kuṭha- 'Costus speciosus'. No other Indo-Aryan dialect has kept the -ṛ- of this word (Turner 1966:174a), except for the Dardic dialects and probably also Gandhāra (see von Hinüber 2001, §231). As we have seen earlier, the Proto-Tocharian transition *u > *a continued to operate in Tocharian B. The solution might be that a dialectal variant of Skt. kuṭha- is the origin of Toch.B kāṣu.

3) Yet another source may be found in (Buddhist) Sanskrit. Most early Sanskrit borrowings show the effect of a set of sound changes characteristic to Tocharian A or B. The distinct changes are specifically
loss of final syllables and devoicing of stops in all positions, as found in, for example, *pilamāti, pilamātii* 'center of the fruit of the Aegle marmelos', borrowed from Skt. *bīla-madhya* (MW:732a). The devoicing/deaspiration of stops (*d, *g* > *t; *g, *q* > *k; *b, *y* > *p, cf. Ringe 1996 §3) is a prototypical Tocharian sound change, which affects Indo-European etymologies as well as early borrowings, like those mentioned above. It is worth noticing that this devoicing/deaspiration was still operating after the split of Tocharian A and B, and the cessation of this sound change can be observed in the Sanskrit borrowings. Later borrowings than those mentioned above keep the voiced stop in initial position, i.e., Toch.B bhakottār ‘portion’ from Skt. bhāgottāra-. At the next stage, the Sanskrit stops are kept except in final position only, i.e., Toch.B bhant ‘Rubia cordifolia’ from Skt. bhānti, Toch.B bhārk ‘Clerodendron spp.’ from Skt. bhāṛg-. The most recent stage keeps the Sanskrit stops in all positions.

- Toch.A kantī ‘gong, cymbal’ translates Skt. *śamyā- ‘stick, staff, wooden pin or peg’ (MW:1054b) in the text A 359.18 | śamyāpraśo niragadā | kantī koṣṭhane yo prā(maṇi talke)|’the beating of a kantī and brahmanical sacrifice’. Skt. śamyāprāsā means ‘wedge-throw’ and is the name of a brahmanical sacrifice (Edgerton 1953:523b). Obviously we have to do with a misinterpretation by the Tocharian translator: Skt. *samyā- also means ‘a kind of cymbal’ (Edgerton 1953:523b; MW:1054b). Toch.A kantī is then an early borrowing from Skt. ganḍī ‘gong’ (Edgerton 1953:208b)\(^4\)

### 4.0 Discussion

The purpose of this overview has been to collect and sum up some linguistic evidence for the reconstruction of the position of Pre-, Proto-, and Common Tocharian in time and space. What conclusions can be drawn from reconstructions of linguistic contacts in pre-historic, para-historic, and historic time, compared to the reconstruction of the language itself and its position within the Indo-European family tree? Starting from the chronologically most recent evidence we might conclude that the two languages, Tocharian A and B, must have separated at a fairly late date. Bactrian as well as Khotanese borrowings should, as we have seen before, already be posited to a Common Tocharian period. If Common Tocharian coincided with a period when Bactrian was influential and also spoken in the Tarim Basin, the time of these transmissions should most likely be put to the beginning of our era. The rendering and adoption of the Buddhist vocabulary is, as we have seen before (3.1.), extended over a period that began in Common Tocharian and then continued separately in Tocharian A and B. The word for Buddha himself was clearly borrowed into Common Tocharian, but other basic Buddhist concepts are different in Tocharian A and B, which indicates that the split of Tocharian A and B happened during the earliest period of adaptation of Buddhism in the Tarim Basin. When is it possible that the Tocharians first heard the word for ‘Buddha’ and decided to borrow and adopt it? According to the description by Haussig (1983:185ff.), the spread of Buddhism in Central Asia was first conveyed via colonizing Indian merchants. This started already during the reign of King Asoka in the 3rd century BC and reached Central Asia, i.e., Parthia and Bactria, first. Somewhat later, probably during the 1st century AD, the colonizing Buddhist merchants reached the Tarim Basin, and by the time of the arrival of the Buddhist missionaries, the merchants were already settled. Therefore, considering the linguistic evidence, it is likely that the Tocharians first came in contact with Buddhism somewhere around the beginning of our era, and adopted the word for Buddha as one of their gods. This belongs to the Common Tocharian period. Somewhat later, with the arrival of the Buddhist missionaries, the process of adopting the religion began. At this time the languages were most likely already separated. The language of the first wave of the Buddhist mission was, on the official and administrative level, a form of Gandhāri. Besides, the Tocharians probably also borrowed words and terms from a local dialect (of early Dardic type?), different from the official language. Later on, Gandhāri became continuously replaced by Buddhist Sanskrit, a process that reflected an increasing sanskritization of the Buddhist administration itself.

Going beyond that, we may assume, with a fairly high degree of confidence, that there was already some contact between Tocharians and Chinese during the Common Tocharian period. Following the idea of a late reconstruction of a split between Tocharian A and B, this exchange of vocabulary could have taken place at any time during the 1st millennium BC. Later, Chinese terms continued to flow in. The assumption of a late split also makes the presence of Turkic borrowings

\(^4\) I thank Seishi Karashima for this parallel.
in Common Tocharian more acceptable, since there might be reason to suspect that ancestors of the Turks might have been present at the eastern edge of the Tarim Basin already at the beginning of our era, which would explain the Turkic borrowings into Common Tocharian. However, this has to be investigated in further detail.

Even earlier than that, in the first half of the 1st millennium BC, we might trace contact between Indo-Iranian and Tocharian, a contact which also continued all the way up to the time of the attested Tocharian language. As concerns language contact preceding that, we are on much more uncertain ground, but some indications, such as the co-occurrence of words in Tocharian and Indo-Aryan (probably also Chinese) from an unknown donor language makes it possible to assume that the Tocharians spent a relatively long period, at least 1,500 years before the earliest attestation of texts, somewhere in the east.

It seems quite unquestionable that Tocharian was cut off from the Indo-European tree at a very early date; thus it retained a large number of archaisms inherited from the Indo-European mother language. Therefore, most probably being separated from the rest of the Indo-European linguistic continuum, Tocharian developed certain non-Indo-European grammatical and typological traits of its own, like case layering and group inflection, which already belong to a Proto- or Common Tocharian period. Whether these changes are due to contact with non-Indo-European languages remains open.

However, from the arguments given above, we must assume that the Tocharian language, from a period beginning somewhere around our era and several centuries onwards, was changing very rapidly. It would seem unlikely that two languages as different as Tocharian A and B were separated by a couple of hundred years only. This is also contradicted by the fact that Tocharian A and B are almost completely stable linguistically during the period of our attested texts, which covers, at least for Tocharian B, almost five centuries. However, it is not unlikely that Tocharian, because of the incursion or influence of other languages combined with concurrent social revolutions, such as the shift of the religious paradigm, experienced a period of very rapid change during the first centuries AD. This is also somewhat reflected in the vocabulary borrowed from Indo-Aryan. Thus, once Tocharian was established and sanctioned as a language used in the translations of the religious texts, the two dialects would have become fixed and used principally unchanged up to the complete extinction in the 10th-11th centuries AD.

Abbreviations:

References
Carling, Gerd 1999 The Tocharian inflected adverbs and adpositions in relation to the case system. Tocharian and Indo-European Studies 8:95-110.
Couvreur, Walter 1955/56 Bemerkungen zu Pavel Pouchas Thesaurus linguae tocharicae dialecti A. La Nouvelle Clio 7/8, 67-98.
Gippert, Jost  

Gray, Russell D., and Quentin D. Atkinson  

Haussig, Hans Wilhelm  

Isebaert, Lambert  

Ito, Michiharu, and Ken-ichi Takahima  

Klingenschmitt, Gert  

Krause, Wolfgang  


Lévi, Sylvain  

Lubotsky, Alexander  

Lubotsky, Alexander, and Sergei Starostin  

Mair, Victor H. (ed.)  

Mallory, James P., and Douglas Q. Adams  

Masica, Colin  

Pan Wuyun  

Pinault, Georges-Jean  


2000  Linguistic contacts of the Tocharian languages. *Handout Göteborg 26/1*.


Pokorny, Julius  

Poucha, Pavel  

Ringe, Don  


Rix, Helmut, Martin Kümmel, Thomas Zehnder, Reiner Lipp, and Birgitte Schirmer (eds.)  
Sagart, Laurent

Schmidt, Karl Honi

Schmidt, Klaus Töth

Sieg, Emil

Sims-Williams, Nicholas

Tischler, Johan

Thomas, Werner

Turner, R.L.

van Windekens, Albert Jonas

---

**Gerd Carling**

Von Gabain, Anne-Marie
von Hüniger, Oskar
2001 *Das ältere Mittelindische im Überblick*. Wien: Verlag der Österreichischen Akademie der Wissenschaften.

Winter, Werner

Gerd Carling
gerd.carling@ling.lu.se