



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
Main Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2017

Tocharian Dialectology

Widmer, Paul

DOI: <https://doi.org/10.1515/9783110523874-034>

Posted at the Zurich Open Repository and Archive, University of Zurich
ZORA URL: <https://doi.org/10.5167/uzh-140805>
Book Section

Originally published at:

Widmer, Paul (2017). Tocharian Dialectology. In: Joseph, Brian; Klein, Jared; Fritz, Matthias. Hand-
book of Comparative and Historical Indo-European Linguistics. Berlin: De Gruyter, 1389-1396.

DOI: <https://doi.org/10.1515/9783110523874-034>

79. The dialectology of Tocharian

1. Preliminary remarks
2. Geographical approach
3. Diachronic approach
4. Palaeographical evidence
5. Absolute chronology
6. Summary
7. References

1. Preliminary remarks

First of all, it should be pointed out that Tocharian A and Tocharian B are considered separate languages here, deriving from a common ancestor usually called Common Tocharian (German *Gemeintocharisch*, French *tokharien commun*). In his contribution on the interrelationship between Tocharian A and B, Lane calls them dialects (Lane 1966), but he himself holds that the differences between Tocharian A and Tocharian B are greater than between any two languages of, e.g., the Slavic or Germanic branch (Lane 1966: 213). So far, however, no general agreement has been reached concerning this matter (cf. Thomas 1984: 126–127 with references). The least that can be said is that the differences observable within the Tocharian B corpus are much less substantial than those between Tocharian A and Tocharian B. All in all, the two idioms differ from each other too much phonologically, lexically, and esp. morphologically to be treated as dialects of a single language (cf., e.g., Winter 1998: 155).

There is virtually no phonological or morphological diversity within the Tocharian A corpus of texts that could be described in terms of dialectology; thus, the dialectological description of Tocharian is first and foremost concerned with Tocharian B.

2. Geographical approach

The first linguistic features to attract the attention of scholars were the so-called “*MQ-Schreibungen*” found in texts from western sites around Kucha. These particular spellings clearly diverge from what is usually found in standard Tocharian B texts unearthed in the central region of Tocharian settlement in and around Šorčuq. In his thorough synchronic treatment of the Tocharian B verbal system, Krause listed a series of features that occur systematically in texts found in the western part of the Tocharian B area, especially in the caves of Ming-oi Qizil (MQ) near Kucha and, to a lesser extent, in Kucha itself (Krause 1952: 1 ff.). Krause suggested cautiously that the texts containing MQ-spellings basically belong to an older layer of Tocharian B. According to his interpretation, it would be incorrect to speak of Tocharian dialects at all.

Based on Krause’s description, Winter (1955) gave a systematic account of the distribution of some 25 phonological features in the texts of the Berlin collection that set the MQ texts apart from the texts written in the standard language. Additionally, he isolated a third distinct group of texts, all of which were discovered in the area around Turfan. Since Winter was unable to devise a chronological scenario for the corpus of Tocharian B, he pursued the matter from a purely synchronic point of view. Winter succeeded in

Tab. 79.1: Shibboleth Features of the Western Dialect as Opposed to the Central Dialect

Western Dialect	Central Dialect
occurrence of < <i>ä, a, ā</i> > not linked to stress rules < <i>ō_i</i> > < <i>ē_u, ew</i> > < <i>ō_u, ow, au, auw</i> >	occurrence of < <i>ä, a, ā</i> > linked to stress rules < <i>oy</i> > < <i>au</i> > < <i>au</i> >

Tab. 79.2: Shibboleth Features of the Eastern Dialect as Opposed to the Central Dialect

Eastern Dialect	Central Dialect
/ <i>ä</i> / spelled < <i>i</i> > in palatal environment - <i>ś</i> - <i>w</i> - <i>św</i> -	/ <i>ä</i> / spelled < <i>ä</i> > in palatal environment - <i>c</i> - <i>p</i> - <i>śtw</i> -

isolating three major dialectal groups: the Central Dialect (area of Šorčuq), the Western Dialect (area of Kucha; esp. MQ), and the Eastern Dialect (area of Turfan). The shibboleth features given by Winter (1955: 224) are provided in tables 79.1. and 79.2.

3. Diachronic approach

As mentioned above, Winter's dialectological categorization is based on the premise that all differentiation within Tocharian B should be ascribed to geography rather than chronology. This premise was not challenged until 1977, when Peter Stumpf in his habilitation thesis (published posthumously in 1990) returned to the position upheld by Krause (1952) and argued convincingly that the diversity in Tocharian B can best be accounted for by positing chronological layers within one single language. Methodologically, he extended the scope of the investigation by taking into account morphological features more comprehensively than Winter.

Stumpf isolated two major linguistic layers, I and II, in the manuscripts, with I being the older layer and II the younger layer (Stumpf 1990: 74, 108).

In layer I, Stumpf identified three sublayers on the basis of the spelling rules for the vowel triad *ä, a, ā* (Stumpf 1990: 76–79) and the spelling of the *u*-diphthongs (Stumpf 1990: 79–82; cf. table 79.4).

At this point, Stumpf faced the problem that, in the absolute chronology he established, the oldest manuscripts of Tocharian B date to the middle of the 7th century CE, while the linguistically definitely younger layer II type documents occur as early as the 7th century as well (Stumpf 1990: 154 f.). This overlap in written attestation contradicted Stumpf's concept of two chronologically layered language varieties. He solved the problem by assuming that the older layer I type language represents the standard that was coined during the heyday of the Kuchean kingdom in the 4th century CE and that contin-

Tab. 79.3: Principal Features Used by Stumpf (1990: 64–107)

Feature	Ex. Layer I	Ex. Layer II	
consonant cluster simplification	<i>nest</i>	<i>nes</i>	‘you are’ (2 nd sg.)
assimilation	<i>šarm</i>	<i>šaräm</i>	‘reason’
articulatory simplification	<i>parna</i>	<i>parra</i>	‘outside’
fricativization	<i>plāc</i> [-c]	<i>plās</i> [-ç]	‘speech’ (obl. sg.)
morphological/lexical variation	<i>šalype</i> [-p-]	<i>šalywe</i> [-β-]	‘salve’
	<i>skente</i>	<i>stare</i>	‘they are’
	<i>wes</i>	<i>wesäm</i>	‘us’

Tab. 79.4: Principal Features of Stumpf’s Sublayers

Feature	I A (≈ MQ)	I B	I C (≈ Š)	II (≈ Turfan)	
accented /ä/	<i>pälsko</i>	<i>pälsko</i>	<i>pälsko</i>	<i>pälsko</i>	‘thought’
unaccented /ā/	<i>šärmänmä</i>	<i>šärmänma</i>	<i>šärmänma</i>	<i>šärmänma</i>	‘reasons’ (nom./obl.)
*æu-diphthong	<i>ēurtse(/aurtse)</i>	<i>ēurtse/aurtse</i>	<i>aurtse</i>	<i>aurtse</i>	‘broad’

Tab. 79.5: Correspondences between Winter (1955) and Stumpf (1990) According to Stumpf (1990: 149)

Winter’s Dialects	Stumpf’s Layers
Western Dialect	I A & I B
Central Dialect	I C
Eastern Dialect	II

ued to be used as a literary language for more than 300 years (Stumpf 1990: 156). According to Stumpf, this standardized literary form underwent, in the course of time, successive influences from non-literary registers that continued to evolve. At the stage of layer II documents, the non-standard variety had finally replaced the old formal style (Stumpf 1990: 144–147, 157). Stumpf’s scenario thus combines historical and sociolinguistic arguments.

4. Palaeographical evidence

Until recently, the study of palaeography did not play a decisive role in settling the question. Although the editors of the Tocharian B texts stored in Berlin (Sieg and Siegling 1949, 1953) made sufficiently reliable remarks about the writing ductus of each fragment, their judgment had to remain somewhat vague since no investigation of Tocharian palaeography had been undertaken at the time of their editorial work. Even today, a detailed investigation is still lacking, and it is to be hoped that the ongoing digitalization will at last make available scans of all texts necessary for this task.

Tab. 79.6: Palaeographical Periodization According to Malzahn (2007: 259, 296–297)

Period	Text with Shabboleth Signs
Most Archaic	THT 1520
Middle Archaic	B 273 ff.
Early Common Archaic	THT 1661 ff.
Late Common Archaic	B 133 ff.

Winter, who did not yet have the opportunity to check the palaeographical labeling by Sieg and Siegling, already noticed (Winter 1955: 220–221) that ligature writing is much more frequent in texts of western origin (i.e. from Kucha), whereas *virāma* writing clearly predominates in standard Tocharian B texts from Šorčuq and in texts from the eastern sites around Turfan. This line of orthographic argumentation was integrated into his scenario by Stumpf as well (Stumpf 1990: 105–107). He expanded on it by adding the spelling of older *k_wu_i* vs. younger *kwri* ‘when(ever), if’ and older *šai* vs. younger *šey* ‘was’ (3 sg. imperfect). Incidentally, *šai* vs. younger *šey* (probably *še* [*< šai*] + *y*) belongs to the phonological level, while *ō_i* vs. *oy* (Winter 1955: 217) is an orthographical matter.

Winter must undoubtedly have been aware of the fact that scribal habits do not carry great weight in establishing synchronically defined dialectal areas. Indeed, the parallel development of specific scribal habits with linguistic change as, e.g., consonant cluster simplification, suggests that the orthographic norm developed alongside the language. In this light, it would seem that Stumpf’s chronological approach is superior to the geographical one proposed by Winter.

Additional evidence in support of Stumpf’s diachronic classification has been adduced by Malzahn in a pioneering palaeographical analysis of the Tocharian B texts, which was facilitated by the high resolution scans of the Berlin and London texts now publicly accessible on the world wide web (<http://titus.fkidg1.uni-frankfurt.de/texte/tocharic/thtcat.htm>; <http://idp.bl.uk>).

Malzahn (2007) provides a solid framework for the palaeographical development of the Tocharian B texts by taking as a starting point the oldest type of the local brāhmī script used in Sanskrit manuscripts, viz. the Early Turkestan Brāhmī (alphabet r), and tracing it down to the standard Northern Turkestan Brāhmī in a series of steps.

The palaeographical layering established by Malzahn fits nicely with the linguistic one by Stumpf, so that, at this point, it can hardly be doubted that the linguistic diversity observable in the Tocharian B texts has to be ascribed, first and foremost, to language development over time.

5. Absolute chronology

With regard to absolute chronology, carbon-14 datings of Tocharian manuscripts offer corroborating evidence for a chronological layering. Tamai (2005) has made available the results of a carbon-14 analysis of nine Tocharian B manuscripts (see Table 79.7). The implications of Tamai’s findings are, first of all, that Tocharian B is attested in

Tab. 79.7: C-14 Dating of Tocharian B documents (Tamai 2005)

MS	Location	C-14 Dating
B 333	Ming-oi Kizil (R)	CE 394–473
B 240	Ming-oi Kizil	CE 428–524
B 601	Kucha	CE 669–780
B 352	Qumtura	CE 679–776
B 178	Sāngim	CE 697–716
B 367	Murtuk	CE 737–773
B 483	Ming-oi Kizil	CE 770–888
B 296	Xocho	CE 1178–1255

writing roughly from 400 CE to 1200 CE, thus supplying additional reasons to assume that the diversity of Tocharian B is due at least partly to language change over time.

From this sample of dated texts a clear pattern emerges: The oldest manuscripts (B 333 [MQ], B 240 [MQ]) are from the western area around Kucha, i.e. Winter's western dialect, and all manuscripts from either central Šorčuq or eastern Turfan areas are much younger than the oldest manuscripts. Furthermore, most of the texts written in Winter's central dialect date from the second half of the 7th century onwards.

The c-14 datings also resolve some of the chronological difficulties Stumpf encountered: They show that not all of the manuscripts that display layer I A or I B features necessarily have to have been written (or copied) in the 7th century CE or later in a literary form that was coined three centuries earlier (Stumpf 1990: 156); at least some of them are originals written in the 4th or 5th century CE.

On the basis of the c-14 datings, Adams (2006: 386) convincingly proposes four chronological/geographical stages:

1. Early Tocharian B (ca. 400–600 CE) only in Kucha and environs
2. Middle Tocharian B (ca. 600–900 CE) everywhere in “attestation area”
3. Late Tocharian B (ca. 900–1100 CE) in Turfan and Kucha and environs
4. Very Late Tocharian B (ca. 1100–1300 CE) in Turfan

This scenario obviously relies on a small sample of dated texts, and not every text is necessarily contemporary with the medium it is written on (cf. the remarks on B 558 in Tamai 2005: 4). Additionally, an older literary style can be imitated by a skilled writer, thus obscuring the picture further. However, it can hardly be due to mere coincidence that the absolute dating squares perfectly with the results of the palaeographical analysis performed by Malzahn and the linguistic statements by Stumpf.

Concerning the younger characteristics of the most eastern variety of Tocharian B in Turfan, Stumpf suggests that they were introduced to the written language through strong interference from more informal styles. He reached this conclusion mostly because some features such as consonant cluster simplification are not confined to texts of eastern provenance, but also occur in documents containing profane texts from Kucha and Šorčuq. In Malzahn's assessment (Malzahn 2007: 289 f.), the peculiarities of the eastern texts are such that one should rather treat them as representing a dialect of their own. In the same breath, she admits that the formal language of the Turfan area population may have been influenced by more informal registers to a higher degree than elsewhere and that both scenarios need not be mutually exclusive.

It should be emphasized that the purely chronological scenario as designed by Adams harmonizes perfectly with Stumpf's interference concept. The assumption would be that low register features (and some peculiarities developed through contact with Tocharian A) had already entered the language of the normative literary standard by the time literary text production was being established in Turfan.

6. Summary

The diversity of the Tocharian B corpus may thus well be described in terms of chronology, implying an eastward migration of the Tocharian B scriptorial activities and a successive modernization of the literary standard over time.

Nevertheless, it is by no means ruled out that Tocharian B had different dialects, and a subset of their constitutive features may be identical to the features ascribed to different registers and chronological stages. However, on the basis of the available evidence and the research done so far, their existence is difficult to prove. It would be essential to find more morphological and syntactic evidence in addition to the phonological arguments predominating so far. A relevant syntactic phenomenon may be seen in the use of a PP with the preposition *y* 'in' (otherwise only known as the first member *y[n]/-* of compounds) instead of the usual inflected case form to express locational relationship in three bilingual (Tocharian B and Prākṛit) documents containing commercial records (Schmidt 2001). The contents of these texts present some difficulties, but the correspondences in table 79.8. (Schmidt 2001: 18 f.) are sufficiently clear.

Note that the Prākṛit version of THT 4062 seems to have taken over the PP construction from Tocharian.

Evidence for dialectal diversity in morphology is present in *śemare*, 3 plural preterite of the verb *kām* 'to come'. It is glossed, and thereby clearly labeled as unusual, in the graffito G-Su 35 as *kameṃ*, the standard form (cf. Widmer 2001: 186 f.). But again, it is difficult to ascertain whether these differences should be ascribed to dialectal diversity or to level and/or style of speech.

In summary, the facts suggest that the linguistic diversity in the Tocharian B documents is due to language change over time and to register phenomena, although the existence of dialects, however defined, cannot be ruled out.

Tab. 79.8: Preposition *y* in Three Bilinguals

MS	Toch.	Kucha-Prākṛit	
THT 4059	a1 <i>y lāpar</i> ₂	b1 <i>laparamiṃ</i>	'in Lapar' (<i>Lapār</i> , cf. Stein 1928: 830) SI P/141
THT 4059	a2 <i>[y]/[aika]₂r</i>	b2 <i>laisāram[mi]</i>	'in Laikar'
THT 4062	a1 <i>y tākkaiṃ</i>	b1 f. <i>i [t]ākkai</i>	'in Tākkain'

7. References

- Adams, Douglas Q.
2006 Some Implications of the Carbon-14 Dating of Tocharian Manuscripts. *Journal of Indo-European Studies* 34: 381–389.
- Krause, Wolfgang
1952 *Westtocharische Grammatik*. Band I. *Das Verbum*. Heidelberg: Winter.
- Lane, George Sherman
1966 On the Interrelationship of the Tocharian Dialects. In: Henrik Birnbaum and Jaan Puhvel (eds.), *Ancient Indo-European Dialects*. Berkeley: University of California Press, 213–233.
- Malzahn, Melanie
2007 The Most Archaic Manuscripts of Tocharian B and the Varieties of the Tocharian B Language. In: Melanie Malzahn (ed.), *Instrumenta Tocharica*. Heidelberg: Winter, 255–297.
- Schmidt, Klaus Totila
2001 Entzifferung verschollener Schriften und Sprachen, dargestellt am Beispiel der Kučā-Karoṣṭhī und des Kučā-Präkritis. *Göttinger Beiträge zur Asienforschung* 1: 7–35.
- Sieg, Emil and Wilhelm Siegling
1949 *Tocharische Sprachreste, Sprache B, Heft 1: Die Udānalankāra- Fragmente, Text, Übersetzung und Glossar*. Göttingen: Vandenhoeck & Ruprecht.
- Sieg, Emil and Wilhelm Siegling
1953 *Tocharische Sprachreste, Sprache B, Heft 2: Fragmente Nr. 71–633*, aus dem Nachlaß herausgegeben von Werner Thomas. Göttingen: Vandenhoeck & Ruprecht.
- Stein, M. Aurel
1928 *Innermost Asia, Detailed Report of Explorations in Central Asia, Kan-su and Eastern Īrān*, Vol. II. Text. Oxford: Clarendon Press.
- Stumpf, Peter
1990 *Die Erscheinungsformen des Westtocharischen. Ihre Beziehungen zueinander und ihre Funktionen*. (TIES Supplementary Series 2). Reykjavík: Málvísindastofnun Háskóla Íslands.
- Tamai, Tatsushi
2005 Paläographische Untersuchung und ¹⁴C-Prüfung. Digitalisierung der chinesischen, tibetischen, syrischen und Sanskrit-Texte der Berliner Turfansammlung, 2.6. 2005 – Berlin. <http://www.bbaw.de/bbaw/Forschung/Forschungsprojekte/turfanforschung/bilder/Tamai.pdf> [Last accessed 21 October 2006].
- Thomas, Werner
1984 *Die Erforschung des Tocharischen (1960–1984)*. Stuttgart: Steiner.
- Widmer, Paul
2001 Ererbtes und Geneuertes: Die Präteritumklasse VI des Tocharischen. *Studia Etymologica Cracoviensia* 6: 181–191.
- Winter, Werner
1955 A Linguistic Classification of “Tocharian” B Texts. *Journal of the American Oriental Society* 75: 216–225.
- Winter, Werner
1998 Tocharian. In: Anna Giacalone Ramat and Paolo Ramat (eds.), *The Indo-European Languages*. London: Routledge.

Paul Widmer, Zurich (Switzerland)

