

# T Tocharian: An Indo-European language from China.

Tocharian is a language that was spoken in the Tarim Basin in the Northwest of present-day China (Xinjiang region, north of Tibet). In the middle of the Tarim Basin there is a large desert, which is surrounded by several oases and enclosed by high mountain ranges. Tocharian is an Indo-European language, related to Latin, Greek, Celtic, and, among many others, English. A few examples suffice to illustrate this: *mātār* ‘mother’; *pātār* ‘father’; *protār* ‘brother’; *ñem* ‘name’; *ṣkas* ‘six’; *ke<sub>u</sub>* ‘cow’.

## Michaël Peyrot

studied Comparative Indo-European Linguistics and Dutch Language and Literature at Leiden University, where he also defended his PhD thesis *The Tocharian subjunctive* in 2010 (published 2013 with Brill). From 2011 to 2014, he worked at the University of Vienna for *A comprehensive edition of Tocharian manuscripts*. He then moved to Berlin for his Marie Curie project *Niya Tocharian: language contact and prehistory on the Silk Road* (2014–2016) at the Berlin-Brandenburg Academy of Sciences and Humanities. His NWO-funded VIDI project *Tracking the Tocharians from Europe to China: a linguistic reconstruction* at Leiden University runs from 2016 to 2021.

Today, the Tocharian language is extinct. How is it known altogether? It is attested in paper manuscripts that have been found on the northern edge of the Tarim Basin, in the territory of the former city-states Kuča, Yānqí and Turfan. These manuscripts, dating from 500–1000 BCE, could be preserved until the present day, thanks to the extremely arid desert climate in the region. Nevertheless, the pieces that survive are only fragments of a Tocharian literature that must once have been quite substantial. The number of manuscript fragments can be estimated at 9,000 for variety “B”, originally from Kuča, but also found in Yānqí and Turfan, and 2,000 for variety “A”, originally from Yānqí, but also found in Turfan. However, these are mainly small

pieces of larger leaves: the number of leaves that are completely preserved is only a couple of hundred, and these are mostly just a single leaf of a larger text.

In order to decipher the content of the fragmentary manuscripts, better-preserved parallels in other languages are crucial. Fortunately, these do in many cases exist: Tocharian literature is almost entirely Buddhist. Buddhism arose in what is today northern India and Nepal, in the 6<sup>th</sup> century BCE. When emperor Aśoka, who reigned over almost the entire Indian subcontinent, made Buddhism the state religion in the 3<sup>rd</sup> century BCE, it spread far beyond its place of origin. From Gandhāra in present-day northern Pakistan, it then expanded



northwest into Afghanistan, where it flourished in the Kushan empire, as well as north into the Tarim Basin, from where it spread further into central China. The fact that anything is known at all about Tocharian is due completely to the spread of Buddhism into the Tarim Basin. Not only can the texts be deciphered thanks to parallel texts in other languages, Buddhism was also the reason why Tocharian and several other languages of the region were written down in the first

In order to decipher the content of the fragmentary manuscripts, better-preserved parallels in other languages are crucial.

place. Initially, Buddhist literature was not written in the local languages, but only in the Middle Indian language of Gandhāra, Gāndhārī. The transmission of the texts must also, to a large part, have been oral. From the middle of the first millennium onwards, texts were written down in the local vernaculars. These were Tocharian A and B in the northeast of the Tarim Basin, the Iranian language Khotanese in the southwest of the Tarim Basin, and later also Tumšūqese, related to Khotanese, in the northwest. All four languages are written in a variety of Brāhmī, a family of Indian scripts. Parallel to texts in the local languages, Sanskrit Buddhist texts were produced, as Sanskrit had replaced Gāndhārī as the language of Buddhism in the region.

With all Tocharian Buddhist literature set in India, it comes as no surprise that the Tocharian language contains many words that are borrowed from Sanskrit. Almost the entire lexicon of religious terms is Sanskrit, and in most cases they are easily recognisable because they contain letters that otherwise do not occur in native Tocharian words, such as *th*, *d* and *dh*, which must in normal spoken Tocharian all have been pronounced as *t*. Words of this type are e.g. Tocharian B *bodhisātve* ‘bodhisattva’ (an enlightened being who is to become a Buddha) and *brāhmaṇe* ‘brahmin’ (a member of the class of priests). Only some of the basic religious concepts are expressed with indigenous terms, such as *pelaikne* ‘law’ (Sanskrit *dharma*) and *yāmor* ‘act, fate’ (Sanskrit *karma*). Some words cannot come from Sanskrit, but point to a Gāndhārī source. These were apparently borrowed before Sanskrit became dominant. An example is *ṣamāne* ‘monk’, which goes back to Gāndhārī *ṣamana*, not to Sanskrit *śramaṇa*.

Before the arrival of Buddhism and Indian culture, Tocharian was also influenced by other languages. The most important among these were Iranian. Iranian is a large language family that does not only comprise the Farsi language of Iran, but also, among others, Kurdish, Ossetic in the Caucasus, Pashto in Afghanistan and Pakistan, and smaller languages in Afghanistan, Tajikistan and western China. Some of the Iranian influence in Tocharian can be attributed to its two Iranian neighbours in the Tarim Basin: Khotanese in the southwest and Tumšūqese in the northwest. However, most must derive from several other Iranian varieties. Among these, a small group of words stands

out because they derive from an archaic form of Iranian and point to contacts in the 1<sup>st</sup> millennium BCE, long before the attestation of Tocharian. An example is Tocharian B *etswe* ‘mule’, which has been borrowed from Old Iranian *\*atswa-* ‘horse’, the source of e.g. Avestan (the language of Zoroaster) *aspa-* and Farsi *asb*. The Tocharian B word cannot be from Khotanese or Tumšūqese because the Khotanese word is *aśśa-*, whose *śś* could not have given Tocharian *tsw*.

Even though Tocharian is so heavily influenced by Sanskrit, Gāndhārī, and several Iranian languages, it is not

close at all to them within the Indo-European language family. This is shown, for instance, by the word for ‘horse’, which can be reconstructed as *\*h<sub>2</sub>ek<sub>2</sub>uo-* (cf. Latin *equus*, Greek *híppos*). In Indian and Iranian, which together form the Indo-Iranian branch, the sound *\*k* is reflected as an *s*-sound: Avestan *aspa-*, Sanskrit *áśva-*, Khotanese *aśśa-*. However, in Tocharian it is reflected as a *k*: the inherited word for ‘horse’ is *yakwe* in Tocharian B. The common ancestor of the Indo-European languages, Proto-Indo-European, was spoken in the Eastern European steppe, probably approximately from 4500 to 3500 BCE. There is increasing consensus

that Indo-Iranian, together with other branches of Indo-European, descends from an Indo-European culture called Yamnaya, dated approximately from 3500 to 2500 BCE. From the Eastern European steppe, the Indo-Iranians moved east and then south through present-day Turkmenistan. The Indians moved southeast into India, while the Iranians remained to their north and moved west and east, into Iran and onto the Eurasian steppe.

With Indians south of the Tarim Basin, Iranians in the west of the Tarim Basin, and probably still more Iranians on the Kazakh steppe and possibly even north and partly east of the



Languages of the Tarim Basin around 500 CE (@ Michaël Peyrot)



Even though  
Tocharian is so  
heavily influenced  
by Sanskrit, Gāndhārī,  
and several Iranian  
languages, it is  
not close at all to  
them within the  
Indo-European  
language family.

Tarim Basin, it is highly remarkable that Tocharian does not show any closer resemblance to the Indo-Iranian languages: all influence, even though some is early, is from a later date. At present, the best explanation for this situation seems to be that the Tocharians moved east over the steppe before the Indo-Iranians started to spread. At the eastern end of the steppe, north of the Altai mountains, an archaeological culture is found that is termed “Afanas’evo”. This culture, close to and largely contemporary with Yamnaya (also 3500–2500 BCE), is often thought to represent a very early phase in the development of the Tocharians. Assuming that the Afanas’evo people, who have left no trace of their language, were early Tocharians, the main problem remaining is the enormous time gap of 3,000 years between the end of the Afanas’evo Culture and the attestation of the earliest manuscripts.

Possibly, the link between the Afanas’evo Culture and the Tarim Basin is formed by the so-called Tarim Mummies. The Tarim Mummies are not real mummies, but rather ancient humans that are surprisingly well preserved, due to the extremely arid and in winter very cold climate of the Tarim Basin. They are from several sites throughout the Tarim Basin, and from different periods. Most interesting are the oldest, which date from the early 2<sup>nd</sup> millennium BCE. They belong to the “Xiaohe Horizon”, which comprises the sites of Gümügōu / Qäwrigul, Xiāohé / Ördök and Ayala Mazar, all of which are today in uninhabitable parts of the desert.

Chronologically, it makes perfect sense to connect the early Tarim Mummies with the Afanas’evo Culture on the one hand and with the Tocharian city-states on the other. However, there is no way to be certain of the language of either the Afanas’evo people or the Tarim Mummies given the total absence of written sources. But we can try to reconstruct the migration route of the Tocharians in order to see whether it is possible that Tocharian was spoken in the Tarim Basin already in the early 2<sup>nd</sup> millennium BCE.

In the NWO-funded VIDI project *Tracking the Tocharians from Europe to China* such a reconstruction is carried out based on linguistic evidence. The many layers of contact for which there is evidence in the Tocharian language will be used to establish where and when the Tocharians have been in contact with which other languages. —

The fact that anything is known at all about Tocharian is due completely to the spread of Buddhism into the Tarim Basin.

Phrygians – Tocharian  
– Baleful signs –  
Ebola – The Islamic  
Empire – The temple  
of Kellis – Buddhism  
in Gandhara – The  
Lost City of Salt –  
The Udruh Project

# A spects of globalisation

Mobility, exchange and the development  
of multi-cultural states

# Contents

Mobility and language



**Alwin  
Kloekhorst**

**In the footsteps of  
the Phrygians.**



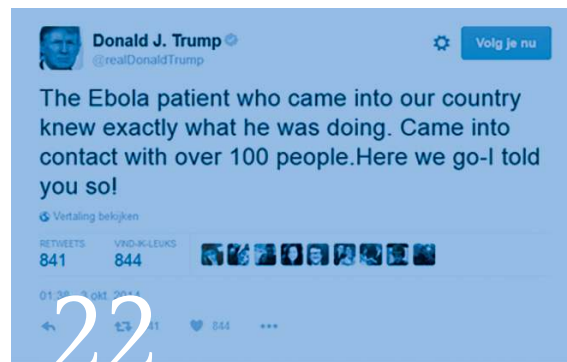
**Michaël Peyrot**

**Tocharian:  
An Indo-European  
language from China.**



**Willemijn Waal**

**In search of the baleful signs.**



**Sara Polak**

**Ebola in the American  
Imagination.**



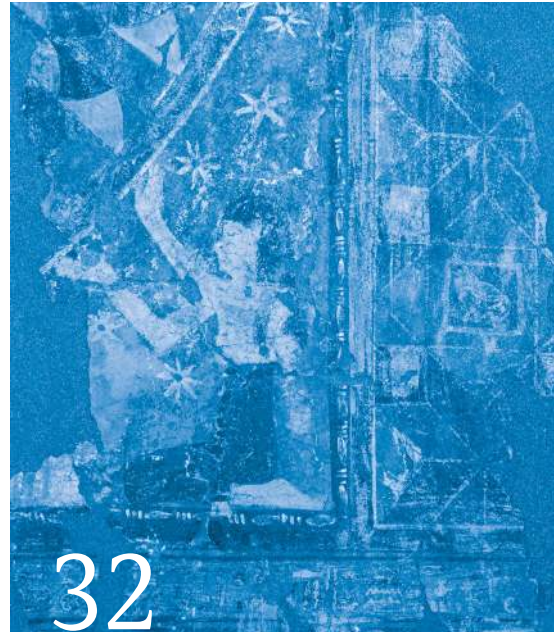
**The multi-cultural state**



26

**Petra M. Sijpesteijn**

The success of the Islamic Empire.



32

**Olaf E. Kaper**

The temple of Kellis at the crossroads between East and West in the Roman Empire.

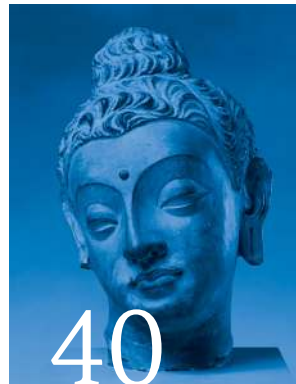
**Trade routes and faiths**



36

**Mark Driessen**

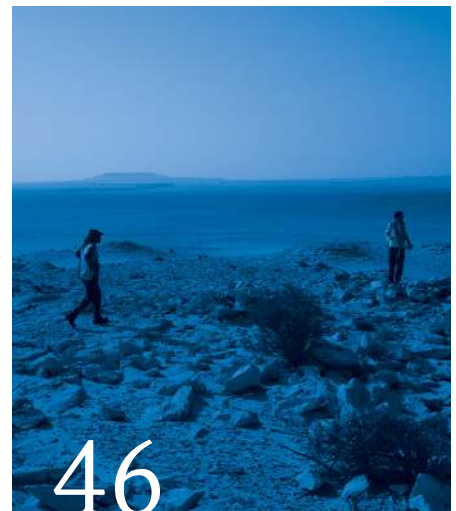
Trade-routes through the steppe.



40

**Marike van Aerde**

Buddhism in Gandhara and beyond.



46

**Ahmad Al-Jallad**

Searching for Ancient Arabia's Lost City of Salt.

# Credits

**New research in the humanities  
Leiden university**

**Edited by**

J. M. Kelder, S.P.L de Jong, A. Mouret

**Photography:**

Rob Overmeer

**Design:**

Just, Leiden

**Print:**

Puntgaaf drukwerk, Leiden

**A Luris publication.  
Leiden 2017.**

Copyright illustration p. 16: "Binghua Wang:  
"The ancient corpses of Xinjiang. The peoples of  
ancient Xinjiang and their culture". Xinjiang 1999.  
Copyright illustration of Papyrus G39726 (page 31)  
Österreichische Nationalbibliothek.

The copyright of all other illustrations and texts rests  
with the various authors.

The Publisher has endeavoured to settle image rights  
in accordance with legal requirements. Any party  
who nevertheless deems they have claim to certain  
rights may apply to the Publisher.

The editors wish to thank Dr. C. Kreuzaler for her  
kind help and the Österreichische Nationalbibliothek  
for permission to reproduce the illustration of Papyrus  
G39726.



Universiteit  
Leiden

LURIS 