

Achtung!

Dies ist eine Internet-Sonderausgabe des Aufsatzes
„Digitization of Tocharian Manuscripts. Short notice about a new project“
von Jost Gippert (1997).

Sie sollte nicht zitiert werden. Zitate sind der Originalausgabe in
Tocharian and Indo-European Studies 7, 1997, 265-266
zu entnehmen.

Attention!

This is a special internet edition of the article
“Digitization of Tocharian Manuscripts. Short notice about a new project”
by Jost Gippert (1997).

It should not be quoted as such. For quotations, please refer to the original edition in
Tocharian and Indo-European Studies 7, 1997, 265-266.

Alle Rechte vorbehalten / All rights reserved:

Jost Gippert, Frankfurt 1998-2011

Digitization of Tocharian Manuscripts

Short notice about a new project

Within the conference dedicated to the fulfilment of "100 Years of Tocharian Studies" (Saarbrücken, Oct. 13-15, 1995), the present state of the Tocharian manuscripts that have been preserved in European museums and collections, and their availability for future research were discussed in a panel discussion. The speakers and the auditory agreed that a project of digitizing the manuscripts should be envisaged as soon as possible, in order both to preserve the data they contain for eternity and to make them more easily accessible to the scholarly world.

A first step in the direction of this aim has meanwhile been undertaken. In a joint effort, the Berlin-Brandenburgische Akademie der Wissenschaften, the Staatsbibliothek Berlin, the Institut für Vergleichende Sprachwissenschaft of the University of Frankfurt, and the Tamai Foundation have started digitizing the Tocharian manuscripts that are preserved in the institutions of the Stiftung Preußischer Kulturbesitz in Berlin. After a test phase which was finished during 1996, the following procedure has been established in accordance with the preservational necessities:

a) Every document is first photographed in its present state. In order to achieve a maximum of quality, this is done using high-resolution colour slide films. This has a double advantage as against using digital cameras: First, the slides can be stored as secondary reference copies of the documents themselves, and second, digitization from slides still gives best results with respect to orthochromaticity and resolution.

b) During the test phase, several attempts were made as to the arrangement of the manuscripts for the photographing. Given that most of the documents are stored in glass frames and should not be removed from them, a suitable background had to be found. It turned out that a bright-coloured paper (white or grey) is best for the purpose, yielding an optimum of contrast with both the manuscript paper and the ink used for writing.

c) The colour slides thus produced are digitized using a high resolution colour slide scanner. With a view to the different purposes the digitized images will be used for, this is done in at least two ways: First, every document is scanned in its entirety, i.e. including the glass frame, with a medium resolution of between 1000 and 1300 dpi (dots per inch). This resolution gives a digital image that fills a normal computer screen, the text being well readable without further enlargement. These images will be made accessible to the public via CD-ROMs and/or the internet in future times; a set of specimens will be available soon under the following World Wide Web address:

<http://titus.uni-frankfurt.de/texte/tocharic>

Second, the individual manuscripts are scanned with a high resolution of 2700 dpi. Using this resolution, a maximum of information can be stored in the digital files, thus meeting the requirements of an eternal preservation of data. Given that this procedure yields enormous file sizes (up to 26 MB per image

— today's CD-ROMs cannot contain more than 650 MB!), a huge amount of storage capacity is necessary for these images. This can be reduced by applying data compression routines, e.g. of the so-called "JPG" format. As this brings about a certain loss of information (the standard rate is 15%), the final decision of the format to store the data in is still being discussed.

d) Many documents are hardly readable, either because of damages or because the ink has faded. Such documents are scanned both as-is, i.e. in the way they appear to the eye, and with the aid of enhancing procedures such as increasing of contrast, intensifying of (ink) colour, etc. Similar procedures can be applied after digitizing too (i.e., using a photo editing software), and it depends on the actual document what steps are necessary for achieving a maximum of readability.

By today, about one third of the Tocharian documents preserved in Berlin have been digitized in the way indicated, and we expect that the photographing and scanning will be finished by the end of 1997. As a first exploitation of the material thus produced, the texts are now being re-transliterated by Tatsushi Tamai in cooperation with Klaus T. Schmidt and J. Gippert in order to establish a basis for a palaeographic investigation.

It is to be hoped that in due time, other institutions that own Tocharian manuscripts will join our efforts to prepare the documents for scholarly analysis and eternal preservation.

Frankfurt, 15.4.1997

Jost Gippert