

Digitization of Turfan Manuscripts

1. Starting point: Conference dedicated to the fulfilment of "100 Years of Tocharian Studies" (Saarbrücken, Oct. 13-15, 1995)
2. Aims of digitizing:
 - 2.1. preserving the data the manuscripts contain for eternity;
 - 2.2. making them more easily accessible to the scholarly world.

3. First step (since autumn 1996):

Digitizing **Tocharian** manuscripts as preserved in the Stiftung Preußischer Kulturbesitz in Berlin (ca. 4300 items): Joint efforts by

Berlin-Brandenburgische Akademie der Wissenschaften

Staatsbibliothek Berlin

Institut für Vergleichende Sprachwissenschaft,
University of Frankfurt

Tamai Foundation

cf. *Tocharian and Indo-European Studies*, vol. 7, 1997

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

- 3.1. Means and procedures

(example: Berlin Tocharian ms. THT 50r):

- 3.1.1. photographing of manuscripts using high-resolution colour slide films (at present: Kodak Ektachrome Professional EPY 5018); advantages as against digital cameras:
 - slides can be stored as reference copies of the documents;
 - best results with respect to orthochromaticity and resolution;

- 3.1.2. choice of suitable background for manuscripts stored in glass frames:

bright-coloured paper (white or grey) to be preferred as against dark-coloured background;

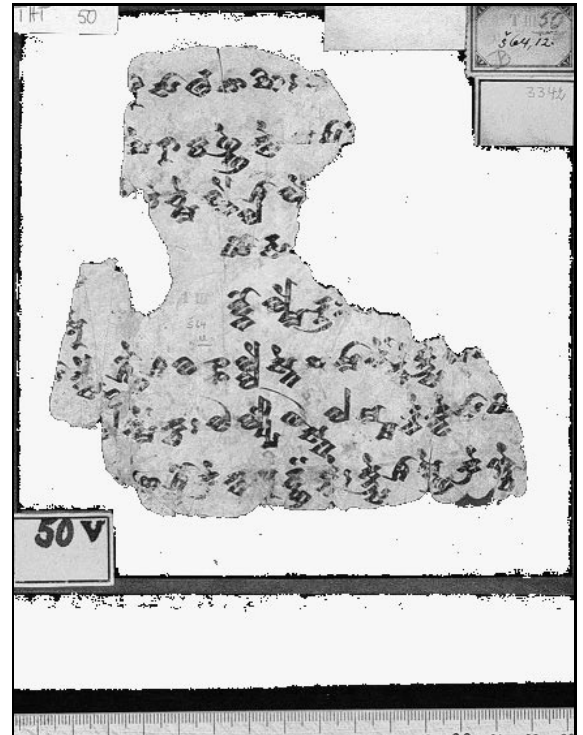
- 3.1.3. procedures of digitizing using high-resolution colour slide scanner (at present: Polaroid SprintScan 32 Plus: 2700 dpi); at least two different scanning processes:

one scanning of the complete picture (manuscript with glass frame, ruler for measuring) with a medium resolution of 1000 to 1300 dpi, yielding digital images that fill a normal computer screen;

one scanning of the manuscript with a maximum resolution (2700 dpi), preserving a maximum of informations for "eternal" storage;

normal settings:

maximum of sharpening ("sharpen detail, smooth noise");
high contrast



THT 50r:

a) total view (650 dpi)



THT 50r:

b) extract, 2700 dpi

- 3.1.4. file sizes (upto 26 MB for a complete colour slide scanned at 2700 dpi) can be reduced to ca. 15% using a JPG data compression format (a minor loss of information has to be acknowledged);
- 3.1.5. for damaged or otherwise hardly readable manuscripts, additional processing may be necessary (contrast enhancing, colour splitting, etc.);
4. Manual work after scanning: re-transliteration of manuscript contents for later linguistic, literary and palaeographical analysis (using "Wordcruncher for Windows" (Brigham Young University / Johnston & Co.; cf. <http://www.wordcruncher.com>):

4.1. Transcriptions and transliteration using Wordcruncher codes:

N.B. In the transliteration, graphical items are separated for palaeographic purposes (index of akṣaras). As against traditional transcription, the so-called "Fremdzeichen" are transcribed as majuscules (e.g., KA instead of *ka*). Conversion into other transcriptional systems remains possible, however.

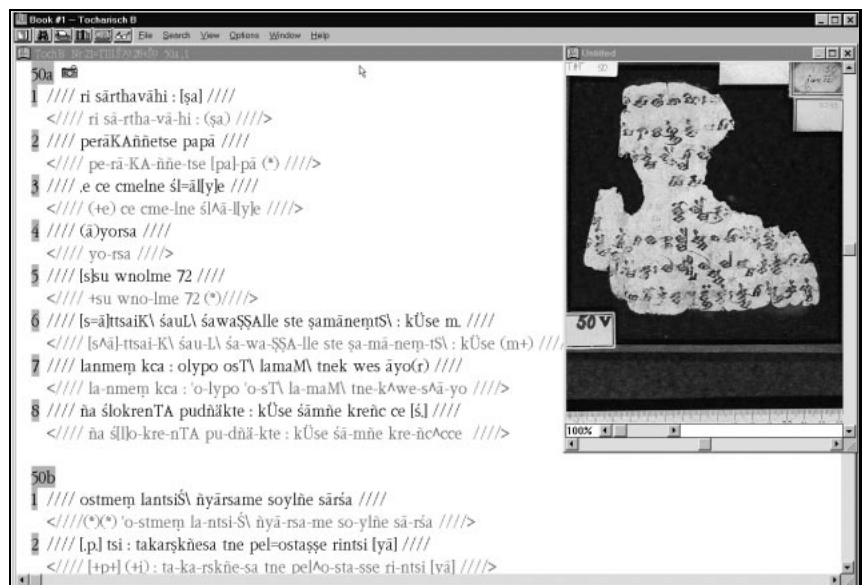
```
lc50a
lp1    /// ri sārthavāhi : [ša] ///
      </// ri sārtha-vā-hi : (ša) ///>
lp2    /// perākAññetse papā ///
      </// pe-rā-KA-ññe-tse [pa]-pā (*) ///>
lp3    /// .e ce cmeľne śl=āl[y]e ///
      </// (+e) ce cme-lne śl^ā-l[y]e ///>
lp4    /// (ā)yorsa ///
      </// yo-rsa ///>
lp5    /// [s]su wnoľme 72 ///
      </// +su wno-lme 72 (*)///>
lp6    /// [s=ā]ttsaiK\ śauL\ śawaššAlle ste šamāneṃtS\ : kuše m. ///
      </// [s^ā]-ttsai-K\ śau-L\ śa-wa-ššA-lle ste ša-mā-neṃ-tS\ : kuše (m+) ///>
lp7    /// lanmeṃ kca : oľypo osT\ lamaM\ tnek wes āyo(r) ///
      </// la-nmeṃ kca : 'o-ľypo 'o-sT\ la-maM\ tne-k^awe-s^ā-yo ///>
lp8    /// nā slokrenTA pudñākte : kuše šāmñe kreñc ce [ś.] ///
      </// nā ś[l]o-kre-nTA pu-dñā-kte : kuše šā-mñe kre-ñc^cce ///>
```

5. Planned next steps:
- 5.1. storing of data on CD-ROMs for easy exchange;
- 5.2. establishing and developing a WWW server for on-line retrieval of texts and digital images (cf. <http://titus.uni-frankfurt.de/texte/texte.html>)
- 5.3. digitization of manuscripts representing other traditions (Iranian: Sogdian, Parthian, etc.; Turkic: Uygur, etc.)

6. Integration into the "TITUS" text data base ("Thesaurus Indogermanischer Text- und Sprachmaterialien" / "Thesaurus of Indo-European texts and language materials"; primary aims:

Preparing all textual materials as relevant for Indo-European studies (Old Indic / Vedic, Old Iranian, Middle Iranian, Anatolian, Old Germanic, Old Celtic, Tocharian etc.) in electronic form for linguistic and literary analysis (cf. <http://titus.uni-frankfurt.de/texte/texte.htm>); cf.:

- 6.1. Call for contributions: "Die Sprache" 32/2, 1986 [1987], 429;
- 6.2. TITUS. Das Projekt eines indogermanistischen Thesaurus; in: LDV-Forum 12/2, 1995, p. 35-47; cf. <http://titus.uni-frankfurt.de/texte/titusldv.htm>;
- 6.3. TITUS — Von der Keilschrifttafel zur Textdatenbank; in: Forschung Frankfurt, 4/1995, p. 46-56; cf. <http://titus.uni-frankfurt.de/texte/titusff.htm>;
- 6.4. TITUS — Alte und neue Perspektiven eines indogermanistischen Thesaurus; in: SIMA 2, Prague 1997, p. 54-89.



Text in "Wordcruncher" representation