

## The TITUS text data base

1. **"TITUS": "Thesaurus Indogermanischer Text- und Sprachmaterialien"** ("Thesaurus of Indo-European texts and language materials"); cf.:
  - 1.1. Call for contributions: "Die Sprache" 32/2, 1986 [1987], 429;
  - 1.2. TITUS. Das Projekt eines indogermanistischen Thesaurus; in: LDV-Forum 12/2, 1995, p. 35-47; cf. <http://titus.uni-frankfurt.de/texte/titusldv.htm>;
  - 1.3. TITUS — Von der Keilschrifttafel zur Textdatenbank; in: Forschung Frankfurt, 4/1995, p. 46-56; cf. <http://titus.uni-frankfurt.de/texte/titusff.htm>;
  - 1.4. TITUS — Alte und neue Perspektiven eines indogermanistischen Thesaurus; in: SIMA 2, Prague 1997, p. 54-89.
2. 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>).
- 2.1. Linguistic scope (ranging from Indic to Celtic), cf. Table 1:

1. **Indica:** Vedica, Sanscritica, Buddhistica, Palica, Pracritica, Rajasthanica, Hindica, Maledivica
2. **Iranica:** Avestica, Persica, Parthica, Persica Mediae Dialecti siue Pahlavica, Khotano-sacica, Sogdica, Persica Recentoris Dialecti
3. **Anatolica antiqua:** Hethitica, Luvica, Palaica, Lydica, Lycica, Milyica, Pisidica, Carica
4. **Tocharica:** Lingua A, Lingua B
5. **Armeniaca:** Armeniaca Antiqua
6. **Baltica:** Prussica, Lettica, Lithuanica
7. **Slavica:** Palaeoslavica ecclesiastica, Palaeobohemica, Palaeopolonica, Palaeoslovenica, Palaeorusica
8. **Germanica:** Gothica, Nordica, Anglica Antiqua, Frisica Antiqua, Saxonica, Antiqua Teutonorum Dialectus, Media Teutonorum Dialectus
9. **Graeca:** Mycenaea, Homerica, Classicorum Dialectus
10. **Italica:** Oscica, Umbrica, Latina
11. **Celtica:** Hibernica, Cymbrica, Brythonica
12. **Indogermanicarum Dialectorum Reliquiae:** Phrygica, Venetica
13. **Linguae Indogermanicarum Vicinae siue Perindogermanica:** Caucasica, Uralica, Turcica etc.

- 2.2. Corpora completed so far (text entry):
  - Iranian:** Avestan; Old Persian; Parthian; Sogdian; Khotanese (Saka)
  - Anatolian:** Luvian, Palaian, Lydian, Lycian, Milyic, Pisidic, Carian
  - Germanic:** Gothic, Old Saxon, Old English
  - Italic:** Oscan, Umbrian
  - Other I.-E. languages:** Phrygian
- 2.3. Corpora under completion:
  - Indian:** Vedic
  - Tocharian:** Languages A and B
  - Germanic:** Old High German
  - Other I.-E. languages:** Venetic
3. Requirements of scholarly text retrieval to be met by electronic "editions":
  - Formatting, encoding and retrieval using "Wordcruncher for Windows" (Brigham Young University / Johnston & Co.; cf. <http://www.wordcruncher.com>);
  - 3.1. "Searching": preindexation instead of sequential text search: immediate results, no size limits for single texts;
  - 3.2. "synchronization" of related texts (e.g. texts and critical apparatuses, cp. Figure 1; various *versiones*, cp. Figure 2);
  - 3.3. linkage of related texts into "search libraries" (e.g., all Vedic texts, cp. Figure 3);

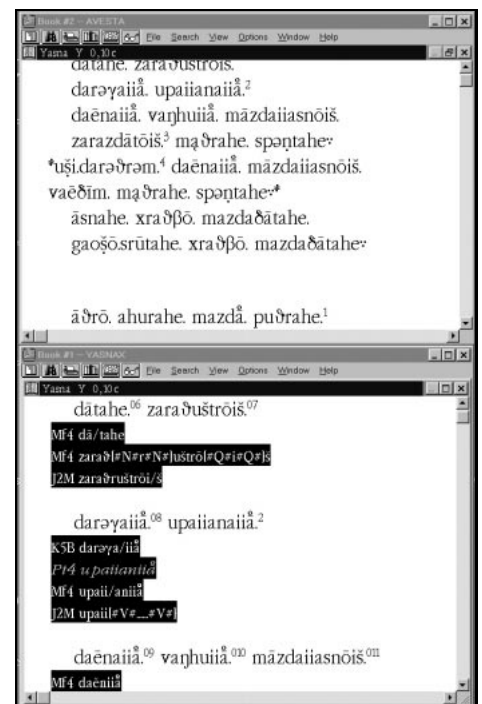


Figure 1: Avestan text with variants

- 3.4. collating various languages and scripts within one text file (important for all kinds of translated literatures, cp. Gothic in Figure 2);
- 3.5. combining editions in original scripts with both "translitative" and "transcriptive" editions (cp. Figure 4);
- 3.6. combining "critically" edited texts with "diplomatically" edited sources: variant readings etc. (cp. Figure 5);
- 3.7. combining textual and graphical data (e.g., manuscript images, cp. Figure 6);
- 3.8. combining textual and analytical data (e.g., linguistic "tags", metrical analyses, cp. Figure 7 and Figure 8);

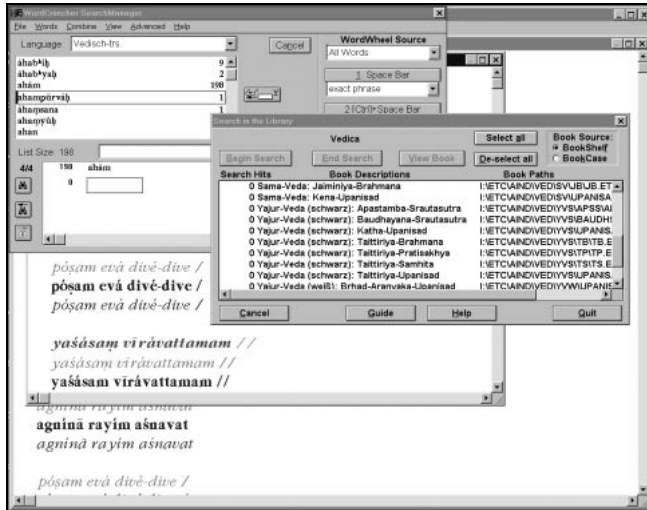


Figure 3: Vedic texts in "Search export library"

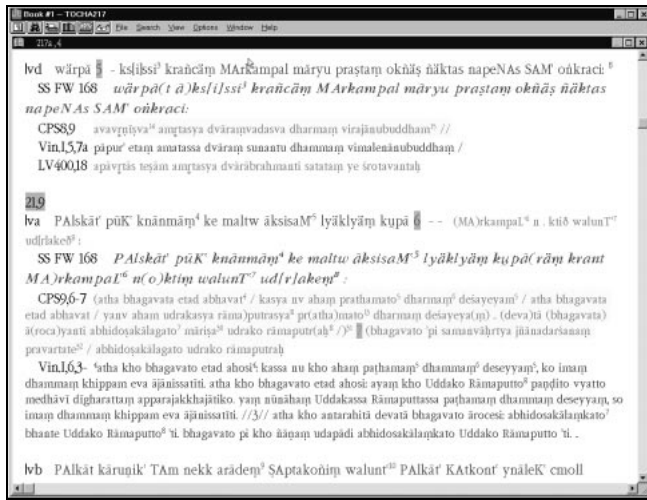


Figure 5: Tocharian (A) text (A 217)

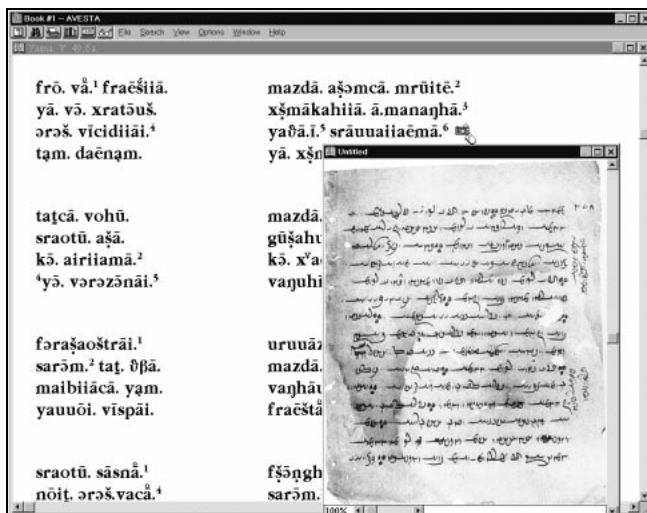


Figure 6: Avestan text with manuscript

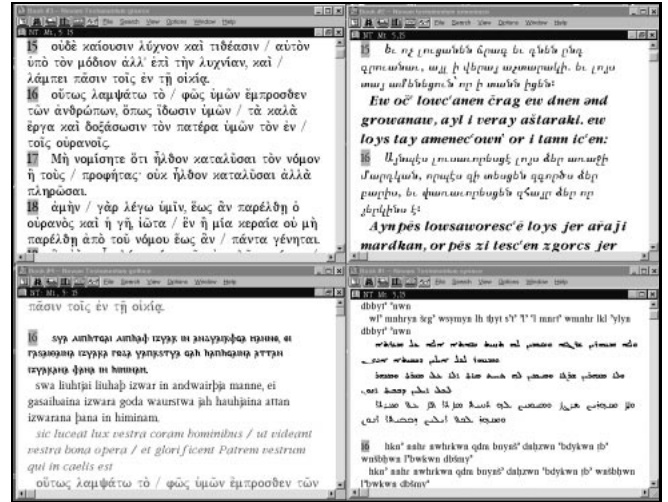


Figure 2: Synopsis of NT versions (Greek, Syriac, Armenian, Gothic, Latin)

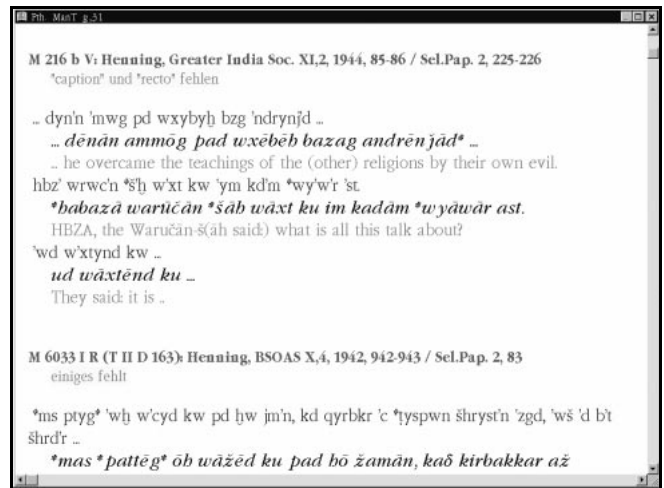


Figure 4: Parthian text (transcription, transliteration, translation)

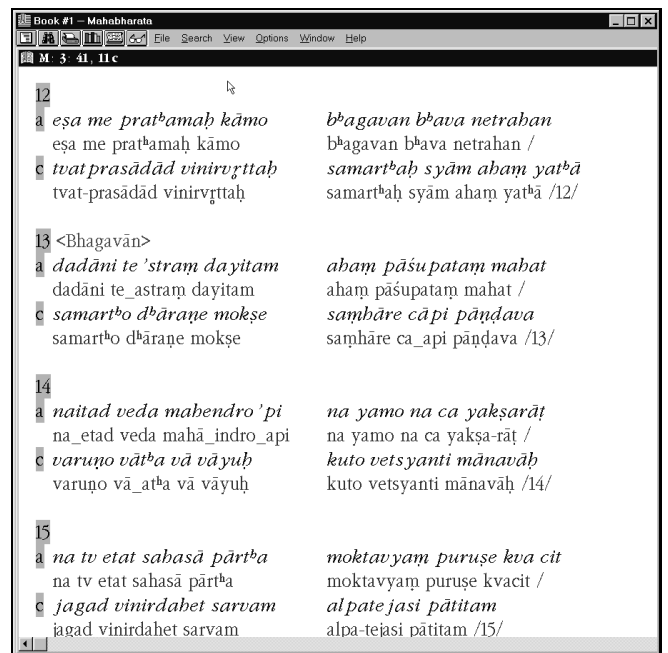


Figure 7: Skt. text with sandhi resolution

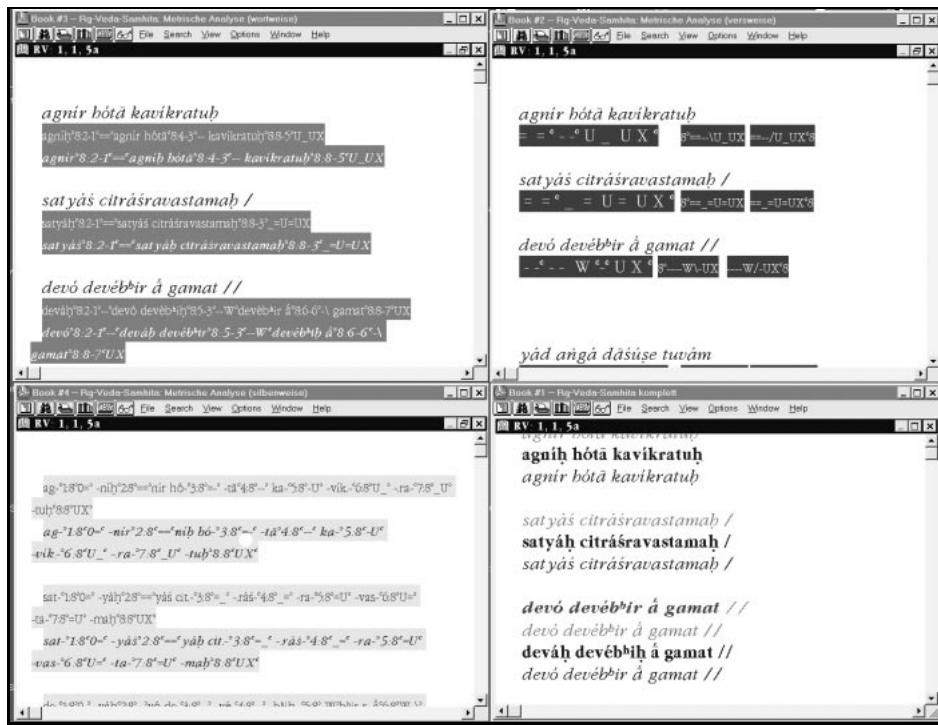


Figure 8:

Rigveda text with analyses

3.9. Preparational activities:

3.9.1. preparation of fonts and keyboard drivers (considering the actual requirements of the texts, cp. Table 2 showing the characters appearing in transcriptional Sogdian, and Figure 9);



Figure 9:

Keyboard definition for Avestan (translitterative)

(!): 2	(0): 3440	(A): 219	(P): 1622	(a): 565	(p): 37882	(ü): 5	(.): 959
("): 13	(1): 10464	(B): 614	(Q): 3	(b): 2126	(q): 5246	(ä): 2	('): 158
(+): 14	(2): 6677	(C): 417	(R): 1555	(c): 12296	(r): 37059	(ñ): 4	("): 174
('): 10829	(3): 4924	(D): 911	(S): 471	(d): 3342	(s): 12573	(ž): 935	('): 87
('): 12031	(4): 4492	(E): 836	(T): 212	(e): 798	(t): 41865	(z): 361	('): 12
('): 12031	(5): 4098	(F): 116	(U): 39	(f): 1465	(u): 147	(:): 1	('): 70
(*): 2177	(6): 3797	(G): 241	(V): 575	(g): 374	(v): 88	(β): 9394	('): 114
(+): 17	(7): 3357	(H): 961	(W): 431	(h): 5627	(w): 45169	(δ): 11192	(:): 60
(,): 454	(8): 3231	(I): 140	(X): 50	(i): 1235	(x): 11931	(γ): 7647	(:): 7
(-): 4272	(9): 3132	(J): 35	(Y): 6556	(j): 374	(y): 57313	(h): 461	(.): 18
(.): 15457	(:): 117	(K): 5354	(Z): 10875	(k): 17513	(z): 6252	(ß): 6	(:): 2259
(/): 770	(;): 78	(L): 1696	([): 10783	(l): 463	('): 943	(h): 87	(:): 42
	('): 138	(M): 1477	(\): 10	(m): 17444	(l): 20632	(š): 11613	(š): 132
	(=): 34	(N): 1835	(]): 10644	(n): 34259	(j): 943	(t): 2831	(:): 2298
	('): 138	(O): 81	(_): 66	(o): 2664		(θ): 831	(:): 344
	(?): 2166		("): 414			(H): 375	
						(Š): 10	
						(Z): 4	

Table 2:

Characters as appearing in the Sogdian (transcriptional) corpus



Figure 10: Wordcruncher "LST-File" (definition of characters)

- 3.9.2. entering of texts (via scanner or keyboard);
- 3.9.3. arranging of text elements (books, chapters, pages, paragraphs, verses, lines, etc.);
- 3.9.4. preparation of language specific character sequence files (LST-files, cp. Figure 10);
- 3.9.5. preparation of text specific format files (SIF-files (cp. Table 3 and Table 4).

```

BC=^^
EC=^_
DOCUMENT_STYLE
  LINE_LENGTH inches 11.0
  LEFT_MARGIN inches 0.1
  RIGHT_MARGIN inches 0.1
  INDEX_OFF superscript subscript
END
CHARACTER_MAP Avestisch
MAPS 1=1 2=2 3=3 4=4 5=5 6=6 7=7 8=8 9=9
MAPS 10=10 11=11 12=12 13=13 14=14 15=15 16=16 17=17 18=18 19=19
MAPS 20=20 21=21 22=22 23=23 24=24 25=25 26=26 27=27 28=28 29=29
MAPS 30=30 31=31 32=32 33=191 34=181 35=182 36=36 37=37 38=38 39=39
MAPS 40=40 41=41 42=42 43=43 44=44 45=45 46=46 47=47 48=48 49=49
MAPS 50=50 51=51 52=52 53=53 54=54 55=55 56=56 57=57 58=58 59=59
MAPS 60=60 61=61 62=62 63=63 64=64 65=166 66=214 67=216 68=219 69=167
MAPS 70=253 71=223 72=183 73=168 74=249 75=221 76=238 77=236 78=240 79=169
MAPS 80=239 81=186 82=82 83=248 84=251 85=170 86=86 87=250 88=184 89=152
MAPS 90=185 91=91 92=92 93=93 94=94 95=95 96=96 97=97 98=98 99=99
MAPS 100=100 101=101 102=102 103=103 104=104 105=105 106=106 107=107 108=108 109=109
MAPS 110=110 111=111 112=112 113=165 114=114 115=115 116=116 117=117 118=118 119=119
MAPS 120=120 121=121 122=122 123=123 124=124 125=125 126=126 127=127 128=128 129=129
MAPS 130=130 131=131 132=132 133=133 134=134 135=135 136=136 137=137 138=138 139=139
MAPS 140=140 141=141 142=142 143=143 144=144 145=145 146=146 147=147 148=148 149=149
MAPS 150=150 151=151 152=152 153=153 154=154 155=155 156=156 157=157 158=158 159=159
MAPS 160=160 161=161 162=162 163=163 164=164 165=165 166=166 167=167 168=168 169=169
MAPS 170=170 171=171 172=172 173=173 174=174 175=175 176=176 177=177 178=178 179=179
MAPS 180=180 181=181 182=182 183=183 184=184 185=185 186=186 187=187 188=188 189=189
MAPS 190=190 191=191 192=192 193=193 194=194 195=195 196=196 197=197 198=198 199=199
MAPS 200=200 201=201 202=202 203=203 204=204 205=205 206=206 207=207 208=208 209=209
MAPS 210=210 211=211 212=212 213=213 214=214 215=215 216=216 217=217 218=218 219=219
MAPS 220=220 221=221 222=222 223=223 224=167 225=225 226=226 227=227 228=228 229=229
MAPS 230=230 231=231 232=232 233=233 234=234 235=235 236=236 237=237 238=238 239=239
MAPS 240=240 241=241 242=242 243=243 244=244 245=245 246=246 247=247 248=248 249=249
MAPS 250=250 251=251 252=167 253=253 254=254 255=255
END
TAB normal
TAB inches .25 .55 .85 1.15 1.45 1.75 2.05 2.35 2.65 2.95 3.25 3.55 3.85 4.15 4.45
TAB_DEFAULT inches 0.3
END
TAB gathas
TAB inches .2 2.8 3.1 3.4 3.7 4 4.3 4.6 4.9 5.2 5.8 6.4 7 7.6 8.2 8.8 9.4 10 10.6
TAB_DEFAULT inches 2.6
END

```

Table 3: Wordcruncher "Structure information file"

```

RULER normal
TAB normal
LEFT_INDENT inches 0.1
RIGHT_INDENT inches 0.1
FIRST_LINE_INDENT inches 0.1
END
RULER gathas
TAB gathas
LEFT_INDENT inches 0.1
RIGHT_INDENT inches 0.1
FIRST_LINE_INDENT inches 0.1
END
PARAGRAPH_STYLE normal
RULER normal
JUSTIFICATION left
END
PARAGRAPH_STYLE gathas
RULER gathas
JUSTIFICATION left
END
LANGUAGE Altavestisch
LANGUAGE_ID Avestan (Old)
CHARACTER_MAP Deutsch
CHARACTER_MAP Avestisch
TEXT_STYLE ial2
LST_FILENAME AVESTTST.ETX
END
LANGUAGE Jungavestisch
LANGUAGE_ID Avestan (Young)
CHARACTER_MAP Deutsch
CHARACTER_MAP Avestisch
TEXT_STYLE ial2
LST_FILENAME AVESTTST.ETX
END
FONT orient
FONT_NAME TITUS-Indoiranisch
FONT_FAMILY roman
CHAR_SET ansi
PITCH proportional
DIRECTION left-to-right
FONT_TYPE TrueType
LANGUAGE Deutsch
END
FONT altavestisch
FONT_NAME TITUS-Indoiranisch
FONT_FAMILY roman
CHAR_SET ansi
PITCH proportional
DIRECTION left-to-right
FONT_TYPE TrueType
LANGUAGE Altavestisch
END
FONT jungavestisch
FONT_NAME TITUS-Indoiranisch
FONT_FAMILY roman
CHAR_SET ansi
PITCH proportional
DIRECTION left-to-right
FONT_TYPE TrueType
LANGUAGE Jungavestisch
END
TEXT_STYLE iiaa22
FONT altavestisch
CHARACTER_STYLE bold
INDEX_FLAG on
FONT_POINTSIZ 22
TEXT_COLOR [0 0 0][255 255 255]
END
TEXT_STYLE iiar12
FONT jungavestisch
CHARACTER_STYLE italic
INDEX_FLAG on
FONT_POINTSIZ 14
TEXT_COLOR [0 0 0][255 255 255]
END
HYPERLINK_STYLE icon
TYPE icon
END
HYPERLINK_STYLE char
TYPE character
HYPERLINK_COLOR [255 0 0][255 255 255]
END

```

Table 4: ("SIF-file")

- 4. Aims for the future:
  - 4.1. transgressing linguistic borders (e.g. with respect to Buddhist texts);
  - 4.2. establishing a server for immediate retrieval (making readable text files downloadable being but an intermediate step; question of formats?).