

CONTENTS

<i>TEXTS AND MANUSCRIPTS: DESCRIPTION AND RESEARCH</i>	3
L. Menshikov. A Fragment of an Unknown <i>Leishu</i> from Tunhuang	3
T. Sultanov. The Structure of Islamic History Book (The Method of Analysis) .	16
<i>TO THE HISTORY OF ORIENTAL TEXTOLOGY.</i>	22
K. Kepping. The Official Name of the Tangut Empire as Reflected in the Native Tangut Texts .	22
<i>PRESENTING THE COLLECTIONS.</i>	33
T. Pang. Rare Manchu Manuscripts from the Collection of the St. Petersburg Branch of the Institute of Oriental Studies, Russian Academy of Sciences	33
<i>ORIENTAL MANUSCRIPTS AND NEW INFORMATION TECHNOLOGIES</i> <i>Correspondence Round table: Arabic/Farsi OCR .</i>	47
A. Matveev. Sakhr Bilingual OCR (Al-Qari' al-Ali). A User's Initial Impressions .	48
J. Bell & P. Zemanek. Test of Two Arabic OCR Programs	55
P. Roochnik. Itisalat OCR Discussion	58
<i>PRESENTING THE MANUSCRIPT</i>	63
O. Akimushkin. <i>Muraqqa'</i> . Album of the Indian and Persian Miniatures of the 16—18th Centuries and the Models of the Persian Calligraphy of the Same Period	63
<i>BOOK AND SOFTWARE REVIEW</i>	68

Color plates: *Muraqqa'*. Album of the Indian and Persian Miniatures of the 16—18th Centuries and the Models of the Persian Calligraphy of the Same Period (see p. 63—67).

Front cover:

Fol. 17a. Portrait of a Man by Riḍā-yi 'Abbāsi, 11.8×8.2 cm.

Back cover:

Plate 1. Fol. 16a. Portrait of Timūr Khān Turkmān by Ṣādiqī beg Afshār, 19.3×11.6 cm.

Plate 2. Fol. 36a. The Darvishes Picnic in the Mountains. Probably Isfahan school, 25.5×14.5 cm.

Plate 3. Fol. 6a. The Shaykh and the Harlot by Muḥammad Yūsuf Muṣavvir, 18.2×11.3 cm.

Plate 4. Fol. 1a. Portrait of Mirzā Jalālā by 'Alī Qulī beg Jabbādār, 16.0×9.1 cm.

RUSSIAN ACADEMY OF SCIENCES
THE INSTITUTE OF ORIENTAL STUDIES
ST. PETERSBURG BRANCH



Manuscripta Orientalia

International Journal for Oriental Manuscript Research

Vol. 1 No. 3 December 1995

THESA
ST. PETERSBURG—HELSINKI

shakids with its numerous priests, a great number of temples and specific way of liturgy, as well as with many every day life restrictions. It influenced greatly other religions. One can find the traces of it in Judaism, Christianity, Islam and Buddhism. At the same time Zoroastrism suffered the influence of other religions. One of the result of it — writing down the Avestan text in the Pahlavi script — should be taken into account.

The translation of such kind of the text could be done by a specialist not only in the field of the history of Zoroastrian religion, but by a historian, ethnologist and philologist with the really encyclopedic knowledge of Indo-European dead and living languages. That is I. M. Steblin-Kamensky who is, fortunately, such kind of a specialist. His Iranian studies are widely known all over the world and we could not miss a chance to congratulate Prof. I. M. Steblin-Kamensky, who has recently been appointed the Dean of the Oriental Faculty of the St. Petersburg University (after Academician M. V. Bogolyubov having retired), on the occasion of his 50th anniversary.

The earliest manuscript of Avesta dates back to 1288. The text of Avesta is still considered to be one of the most difficult ever created in Indo-European languages.

The translation by I. M. Steblin-Kamensky follows not only the sense and exact meaning of the original text, it also reproduces the very structure of the Hymns. Most of the Yashts (Hymns) were created in the form of 7—9 syllable verses (sometimes 4—13 syllable), which was carefully preserved by Prof. Steblin-Kamensky. The poetic style of the translation is very clear and natural. It is absolutely relevant to the ancient Holy Hymns.

The work under review represents an important contribution to the Avestan and Iranian studies in general. Both the common readers and the students of the East will certainly enjoy Prof. Steblin-Kamensky's compact and integrating presentation of the ancient text. The next part of the Avestan translation by Steblin-Kamensky — the Gathas (the Songs of Zaratusht) — is expected in the near future.

F. Abdullaeva

ADOBE ACROBAT 2.0 AS A MEDIUM FOR ELECTRONIC PUBLISHING IN ARABIC STUDIES (WINDOWS VERSION)

Producer: *Adobe Systems Incorporated*

System requirements:

- 386- or 486-based personal computer (or Pentium)
- Microsoft Windows 3.1 or greater
- 4 Mb of RAM for Acrobat Exchange and Acrobat Reader
- 8 Mb of RAM for Acrobat Distiller

Method of testing:

- computer used
Compaq DeskPro XL 590, CPU Pentium 90,
16 Mb RAM
- platform
Microsoft Windows 3.1
Microsoft Windows 3.1 with Arabic support
("bi-l-i'timad")
- texts and PostScript files produced with
Microsoft Word for Windows 6.0a
Microsoft Arabic Word for Windows 6.0a
NISUS 3.4 (Macintosh)

The Adobe Acrobat Pro 2.0 package comes with Adobe Reader 2.0, Adobe Exchange 2.0 (PDFWriter), and Adobe Distiller 2.0, together with a limited purpose version of Adobe Type Manager 3.01. An upgrade of the Reader (version 2.1) is available from Adobe Systems (<ftp://ftp.adobe.com> or <http://www.adobe.com>).

The Windows version of the Adobe Acrobat family of programs performs essentially the same as the Macintosh version described in an earlier review (*Manuscripta Orientalia* 1/2, <http://www.hf-fak.uib.no/i/midtspraak/acrobat.htm>). In general PDF (portable document format) files retain all the formatting and detail of hard copy and can handle both Latin diacritics and Arabic script, although there are occasional problems with screen fonts, and image quality must be reduced if files are to be kept to a manageable size. PDF files can be produced directly from word

processing programs by PDFWriter, which is part of Acrobat Exchange. They may also be generated from PostScript files by Acrobat Distiller, including files containing Arabic fonts. PostScript files can be created from most word processor or page layout programs or by using the printer option offered by Acrobat Distiller Assistant. Acrobat Reader, which allows viewing on screen, simple searching, and printing out, is available free. Acrobat Exchange permits the inclusion in a PDF file of hypertext links to other files, including sound and video clips. Exchange also provides sophisticated searching techniques (see below on Adobe Catalog). Compared to word processor files, PDF files tend to be rather large, but they can be considerably reduced, often to little more than the size of the word processor original, by the exclusion of Type 1 (PostScript) fonts.

Production of PDF Files. It is possible to render Arabic TrueType fonts with PDFWriter on the PC platform using US Word in US Windows and to minimize file size by excluding the fonts from the file, but the Arabic text will not be properly coded on the Macintosh platform. In Arabic Windows, moreover, PDFWriter does not render Arabic script produced by Arabic Word for Windows. Since PDFWriter is therefore generally not appropriate for our purposes, the basic task of interest here is the production of PDF files from PostScript files with Acrobat Distiller. Once a PDF file is produced, Adobe Reader or Exchange, with minor exceptions, can handle it properly.

The way Acrobat Distiller works with TrueType fonts is unfortunately far from satisfactory. TrueType fonts are converted to Type 3 (bitmap) fonts, and cannot be excluded from PDF files as is possible with Type 1 fonts, or with TrueType fonts when using PDFWriter. Files containing Type 3 fonts can be five times, or more, the size of equivalent files based on Type 1 fonts. Moreover, the quality of the screen display of TrueType fonts is inferior to that of Type 1 fonts. Text is sometimes almost illegible, a good result often being obtained only when zooming up to 400% or more. The printout quality is very good, however, being in every respect equal to a printout from the program in which the document was originally produced. The problems with TrueType fonts are particularly significant for the publication of documents containing Arabic script, since all the Arabic fonts in Microsoft Arabic Windows, at least in the version we are using, are TrueType fonts. With Arabic TrueType files from Arabic Windows Common Ground's Digital Paper format gives considerably better results than Acrobat Distiller. The files are smaller, and the screen display is much easier to read.

Another problem seems to lie in the specification of the Arabic TrueType fonts in PostScript files from Microsoft Arabic Windows. The TrueType fonts in the PostScript files are specified only at the end of the file, and interpreting them on occasion seems to cause difficulties for Acrobat Distiller. Quite often they are interpreted as something else (usually Type 1 Courier), or as an offending command. In the latter case no PDF file is created. This problem may at times be connected with combining different sizes of letters.

Sometimes in the case of a longer document (16 pages, 16 point type) only the odd pages were produced in Arabic script, while all the even pages were rendered in Courier. A similar thing occasionally happened when combining Arabic and Latin script, most of the Arabic parts being reproduced normally, while some were rendered as Courier. We found no regular pattern here.

Acrobat Distiller for Windows can be used to make PDF files from PostScript files produced on a Macintosh, and the size of the files can be reduced by eliminating Macintosh Type 1 fonts, including diacritic fonts. It is sufficient to write the names of the fonts in the "Never Embed" list and disable the "Embed All Fonts" option. Unfortunately Arabic fonts cannot be excluded. The results are very good on the screen and excellent when printed out. The Windows version of Common Ground (2.01), which is the only version currently available, is noticeably inferior to Distiller in rendering PostScript files, whatever the original platform.

Displaying of PDF Files. As on the Macintosh, Type 1 fonts must be uninstalled when browsing files that contain them, whether or not they are embedded. The Reader seems to look for the fonts installed in the system and use them for browsing, which can lead to conflicts between them and the fonts in the PDF file. With the exception of the poor screen display of TrueType fonts mentioned above, there are no great problems with browsing in US Windows. In Arabic Windows, however, one must remember to disable Adobe Type Manager, if it is installed, otherwise Acrobat Reader may fail to function, as it did on the machine we were using.

Searching. Acrobat Reader offers non-problematic simple searching ("Find") in lower ASCII Latin text. In Arabic, however, while there are some indications that searching may be possible in a future version of the Reader, this is not the case at the present time. Certain differences between apparently identical words result in the next occurrence of the word not always being found. As in the Macintosh version, there are problems searching for upper ASCII Latin diacritics, and the various Arabic glyphs are not reduced to their least common denominators, which renders meaningful searching almost impossible. Because of the new Acrobat encoding, the keyboard cannot be relied upon to type upper ASCII characters into the Find window. The case is even more complicated with Arabic glyphs. Copying and pasting produces quicker results than the keyboard, but is still not entirely reliable. The simple Find function in Acrobat Exchange performs similarly.

Copying and pasting text. Lower ASCII text may be copied from PDF files opened in Reader or Exchange and pasted into word processor files without difficulty, except that line breaks become paragraph breaks. Copying and pasting of upper ASCII Latin and of Arabic characters gives chaotic results, however. Arabic text originally from Arabic Word for Windows, for example, cannot be pasted back into that program with success.

Adobe catalog. This tool, which includes software licensed from Verity, Inc., and which is sold separately from the Adobe Acrobat Pro package that contains Distiller, Exchange and Reader, creates indexes of the PDF files stored in designated directories. The updating of indexes can be automated and set to run at specific times, every hour, for example. Once the indexes are created, it is possible in Exchange to form various types of queries, including Boolean, stem, and semantically related word searches. Indexing is a very simple and relatively fast process. With a large amount of data, however, updating the index at close intervals could make it difficult to work simultaneously with other programs. Searching for upper ASCII Latin and Arabic, unfortunately, has all the shortcomings mentioned in connection with the Find function in Reader and Exchange. With Arabic, moreover, one must search for glyphs rather than characters, and even words that seem to be identical will not necessarily be found in a search.

Conclusion. Despite the problems mentioned above, Adobe Acrobat remains a superior tool for electronic publishing, whether online or offline. The primary advantage in the Acrobat system seems to be the use of real fonts

which permit real text searches but which do not have to be embedded in all files in a given series. For browsing it is sufficient to open first a single file which has the fonts used embedded. Subsequent files opened do not have to have the fonts embedded and can thus be many times smaller. It is reasonable to assume that the present problems with the encoding of upper ASCII and Arabic glyphs

will be solved with the introduction of a more universal coding system.

This review is preserved in electronic form and in hard copy in the Archive of Electronic Publications of the Section for Middle Eastern Languages and Cultures, University of Bergen (<http://www.hf-fak.uib.no/i/midtspraak/aep.htm>).

J. Bell
