CONTENTS

Front cover:

"Ni Heng (173—198), a poet in the service of Cao Cao". Illustration No. 31 to the Chinese novel *Three Kingdoms* from the Album H-13 preserved in the manuscript collection of the St. Petersburg Branch of the Institute of Oriental Studies (early 19th century), 15.6 × 19.6 cm.

Back cover:

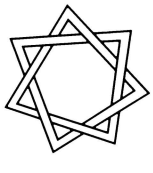
Plate 1. "A high-spirited stone, a divine oriole". Illustration No. 46 to the Chinese novel *A Dream in the Red Chamber* from the same Album, 15.5 × 19.6 cm.
Plate 2. "Shi Ziang-yun falling asleep on the stone bench". Illustration No. 58 to the Chinese novel *A Dream in the Red Chamber* from the same Album, 15.2 × 19.6 cm.
Plate 3. "Lin Dai-yu speaking to a parrot". Illustration No. 57 to the Chinese novel *A Dream in the Red Chamber* from the same Album, 15.5 × 19.5 cm.

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

Manuscripta Orientalia

International Journal for Oriental Manuscript Research

Vol. 3 No. 3 November 1997



TSESA
St.Petersburg-Helsinki

M. Blank, N. Stavisky

CONSERVATION OF MEDIEVAL MANUSCRIPTS IN THE LIBRARY OF THE JEWISH THEOLOGICAL SEMINARY OF AMERICA*

The Library of the Jewish Theological Seminary of America in New York is the repository of one of the great collections of source material for Jewish studies. It includes manuscripts, *genizah* fragments, incunables, sixteenth to twentieth century broadsides, a unique collection of *ketubot* (marriage contracts), *Megillot Esther* (Esther scrolls), archival material, graphics, and rare printed books from the sixteenth century to the present time. The Library is used extensively by international scholars as well as faculty and students of the Seminary.

This paper describes some selected methods and materials used in the conservation of three types of objects in the collection: Maimonides *genizah* fragments, a German thirteenth century *Mahzor* (a holy day prayer book), and a fourteenth century Spanish *Haggadah*. While the techniques described are well known in the United States, they are probably less familiar to practicing conservators in other countries. Though not every stage of the conservation process is described, we hope the selection will be of interest to our European colleagues.

1. Maimonides Material

Over the ages, Jewish communities have followed an established custom whereby worn texts, containing God's name, are not discarded but are gathered in a designated place called a *genizah*, usually prior to collective burial. Over one hundred years ago, the value of such a depository was discovered in the Ben Ezra synagogue in Fustat, old Cairo. The astounding thing was that this hidden collection contained not only sacred texts, but a whole gamut of documents — literary works, poetry, scientific and grammatical texts, philosophical treatises, letters written by both historical personalities and ordinary people, *ketubot* (marriage contracts), commercial inventory records and legal documents, and lost or previously unknown works such as those relating to the Dead Sea Scrolls and *Ecclesiasticus*. These contents shed a new light on a period of Jewish history, particularly in the Islamic world, dating from about the tenth century C.E., about which very little was previously known.

The first significant collector of *genizah* material was a Russian Karaite, Abraham Firkovich, active during the 1860s. His collection was later sold to the Imperial Russian Library in St. Petersburg. In 1897, Professor Solomon Schechter, the primary discoverer of the *genizah*, acquired by far the greatest number of *genizah* documents which are now at Cambridge University. The Seminary collection contains approximately 30,000 fragments [1], the bulk of which were purchased from the great Anglo-Jewish collector Elkan Nathan Adler (1861—1946) in 1923. Adler, an inveterate traveler and lawyer, was among the first laymen to appreciate the significance of the *genizah* fragments he acquired during two trips to Egypt, in 1888 and 1895—6, prior to Schechter's acquisition.

In 1996 and 1997, the Library undertook the conservation of 23 manuscript fragments from manuscripts by Moses Maimonides (1138—1204) and his descendants. Maimonides — who was born in Cordoba, Spain and who died in Fustat, Egypt — was a philosopher, codifier and commentator on Jewish law and texts, a renowned physician, and the seminal figure in Jewish life during the post-Talmudic period. (It is remarkable that when the conservation project had been almost completed, an additional fragment was found and identified.)

All 23 fragments originate from the Cairo *genizah*, with two coming from Schechter's own collection, a gift to the Seminary from his widow Mathilde. This famous letter, signed by Maimonides himself, solicits funds from the

* The authors wish to thank the following people for their help and encouragement: Dr Mayer Rabinowitz, Librarian of The Jewish Theological Seminary of America (J.T.S.L.); Rabbi Jerry Schwarzbard, The Henry R. and Miriam Ripps Schnitzer Librarian for Special Collections, J.T.S.L.; Sharon Lieberman Mintz, Assistant Curator of Jewish Art, J.T.S.L.; Dr Jay Rovner, Manuscript Bibliographer, J.T.S.L.; Evelyn M. Cohen, Assistant Professor of Art History, Stern College for Women, Yeshiva University, New York; Deborah M. Evetts, Book Conservator, The Pierpont Morgan Library, New York; Patricia Reyes, Mellon Conservator, The Pierpont Morgan Library, New York; Dr I. P. Mokretsova and V. Z. Grigorieva, organizers of the Moscow Conference; and Shalom Lipner, for his editing.

community to free Jewish hostages taken captive in the Egyptian town of Bilbays in 1168 by the crusader king of Jerusalem, Amalric. Other items, all from the Adler collection, include a fragment from Maimonides' "Guide for the Perplexed", two draft pages of his *Mishneh Torah,* various responsa, a letter congratulating Maimonides on his second marriage, and various documents and compositions connected with Maimonides and his descendants, or associated with them.

Twenty-two items are on paper; one is on parchment. The sizes vary greatly, from 8.5 × 8.5 cm and 20.5 × 28.9 cm to 42.0 × 14.0 cm, with many variations in between. The paper is generally gray-beige, with no distinct chain or laid lines. An examination of one document with a polarizing microscope indicated that the fiber was hemp. Sheet formation is generally uneven, with clumps of fiber indicating poor beating. Most fragments have been abraded, with numerous tears, cracks, folds and holes. The pH varied between 5.0 and 7.0.

The ink is black, occasionally brownish-black, and appears to be iron-gall. It is generally stable when tested with a drop of deionized water and blotted with filter paper. Very few fragments showed sensitivity to water when tested in this manner. The whole group was soiled with grime and stains of unknown origin, and many appeared to be water stained. Some were repaired (or rather, held together) with crude patches, yellowed paper tape and/or pressure sensitive tape (*figs. 1, 2*).

In the late 1950s, eleven fragments were covered on both sides with silk crepeline. The silk was adhered with a flour paste [2] under strong pressure, and was now a brownish color. The silk often veiled the text, and the strong pressure left an imprint of the textile's weave on the ink itself. The paste changed the texture of the originally soft paper by making the sheet extremely stiff. After consultation with the curators and outside conservation colleagues, it was decided to remove the deteriorating silk.

The preferred technique was to remove the silk crepeline when dry, peeling it off the sheet with the aid of tweezers and scalpels. The disadvantage in this method what that the stiff layer of adhesive remained embedded in the paper, though it was later reduced with damp blotter cleaning. There were also some stubborn documents which did not allow the peeling away of the silk.

A test was made on one document to remove the silk after humidification with damp blotters through Gore-Tex [3]. A sheet of protective lens tissue, followed by the Gore-Tex, the smooth side down, was laid on the object. Damp blotters were placed on top of the Gore-Tex, which was then covered with a Plexiglas sheet lightly weighted to both enhance contact between the object and the moist blotters, and to avoid water evaporation. Frequent inspection ensured the removal of the silk after 10—15 minutes, when the paste was adequately humidified and softened. In one case, while the silk came away easily, ghost writing was visible on the fabric.

When the ink was stable, the fragment was protected with lens paper and damp blotter "washed" between several layers of slightly moist filter paper on recto and verso. A lightly weighted Plexiglas sheet was laid on top. Frequent inspection and changes of soiled filter paper indicated the extraction of stains and degradation products from the paper, and proved to be a very gentle and effective method of cleaning. The paste layer on previously silked items was also reduced, though not completely removed. The paper became more flexible and the inter-fiber bonding strengthened as a result of the humidification process.

Most of the unsilked documents were cleaned in text-free areas with Groomstick [4], a soft, kneadable and malleable eraser primarily composed of vulcanized rubber with a neutral pH. The advantages of gaining access to even very minute areas without touching the letters of text are obvious. Particles of dirt are trapped in the Groomstick and when used with a very light touch, no discernible residue is left. The fragments appeared fresher after the removal of the surface dirt.

Unsightly paper patches were removed with a poultice of 4% methyl cellulose [5] which softened the adhesive, applied over a swatch of thin Japanese tissue or lens paper for easy removal. Sometimes the poultice was mixed with 0.001% amylase enzyme [6]. Paste residue was removed with home-made bamboo and Teflon spatulas. A final local rinse on the suction disc ensured total removal of any remaining adhesive or enzyme.

The suction disc was also useful in the removal of pressure sensitive tape, and particularly stubborn stains.

Tears were usually repaired with Japanese tengujo paper impregnated in our lab with 5%—10% sturgeon glue [7]. Tengujo paper is made of kozo fibers, and is both strong and quite transparent. The paper is laid onto a Mylar [8] support, pasted out with warm isinglass, and left to dry. We cut the repair patch with a scalpel directly on the Mylar which was placed over the document on a light box, taking great care not to pierce the mylar itself. The repair patch was peeled off the Mylar, activated with a tiny bit of dilute paste, laid down over the tear, and then weighted down with a small weight over hollytex [9], blotter and Plexiglas. Holes were repaired with compatible papers, usually tinted with water-color, or with fiber fills, and the document was humidified and flattened between hollytex, blotters, Plexiglas and weights.

Each object, including the one written on a parchment support, was inlaid in a handmade hemp and cotton paper called Akbar (*fig. 2*), made by Griffen Mill specifically for the conservation of Islamic documents [10]. The fragment was laid on a light box and protected by a sheet of mylar, with a sheet of Akbar paper positioned on top. The contours of the object were lightly traced in pencil. After removing the Akbar paper, a sharp scalpel cut away the center shape about a millimeter beyond the pencil tracing. The fragment was placed in the Akbar frame, and attached with thin strips of isinglass paper following the contours of the document and overlapping both the original and the inlay by about a millimeter. Should it ever be necessary to remove the fragment from the inlay, this can be easily accomplished. The inlay technique makes it possible to examine the document without actually touching the original paper.

The Library's curators required certain specifications to be met for the final housing, designed and executed by Deborah Evetts [11]. These were that the fragments be well protected and that, while access to scholars would be assured, no damage to the documents would ensue. The housing was to provide ample protection during possible transportation to other institutions for exhibition purposes while, at the same time, presenting an aesthetically pleasing appearance. A mat assembly, front and back, was constructed from two laminated 4 mil. cream-colored museum quality boards with windows cut out. Mylar film, which is

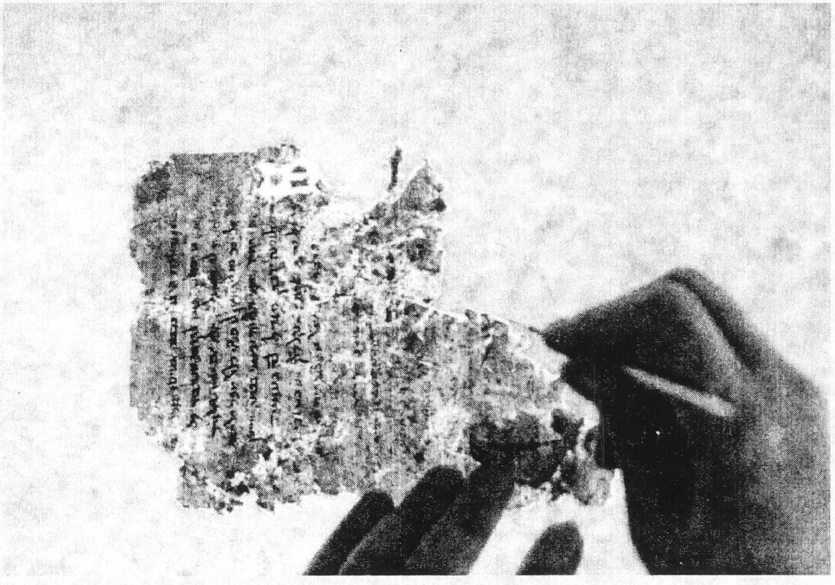


Fig. 2

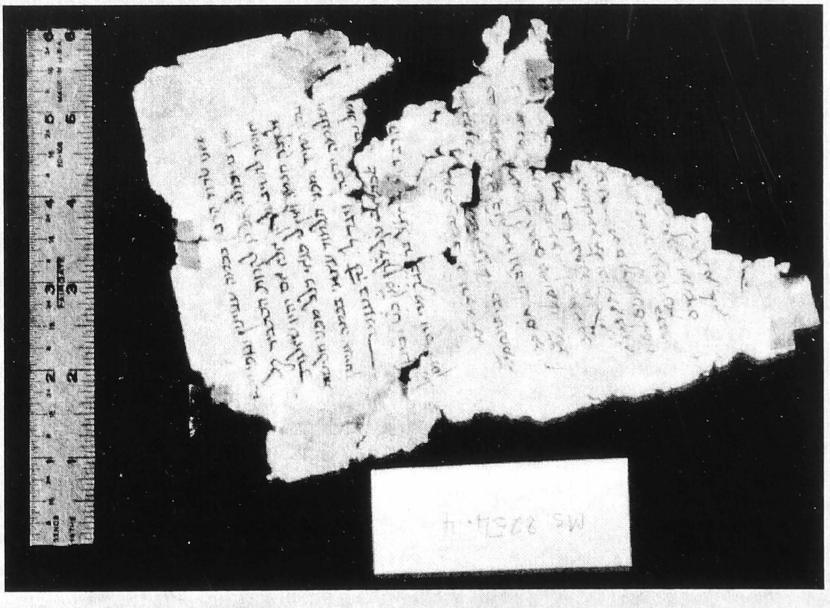


Fig. 1

inert, covers the windows but does not actually touch the object, since it is attached between the two walls of the double board; this space protects the ink from the static electricity of the mylar. Both recto and verso can easily be examined through the windows on either side. The mat is bound into an elegant linen-covered folder, and several folders of the same size are kept in specially made boxes, seven boxes in all.

2. Esslingen *Mahzor*

An interesting story is connected with the identification of the Esslingen *Mahzor*. During a visit to the Bibliotheca Rosenthaliana, Amsterdam, in the summer of 1989, Evelyn M. Cohen, the then Curator of Jewish Art at the Library of the Jewish Theological Seminary of America, viewed a slide presentation including a page of the Esslingen *Mahzor*. Cohen recognized its close resemblance to a manuscript in the Seminary's collection that was lacking a colophon and, indeed, the two volumes — an ocean apart — turned out to be the first and second parts of the same oeuvre. Evelyn M. Cohen and Emile G. L. Schrijver have described the codicology and decoration of the manuscript in great detail [12].

The Esslingen *Mahzor* is a High Holy day prayer book written on calf parchment by the scribe Qalonimus ben Judah in the town of Esslingen, near present day Stuttgart, Germany. The colophon states that the *mahzor* was completed on 28 Tevet 5050 (12 January 1290), which makes it the earliest signed, dated and localized German manuscript known. These types of *mahzorim* were primarily intended for the use of the *hazzan* (cantor) in the synagogue, and generally contained the community's prayers and *piyyutim* (liturgical poems) for the holidays.

Most of the text is written in Ashkenazic square script, though some liturgical instructions and glosses appear in semi-cursive script. The ink is black, occasionally with a reddish-brown halo, and is probably a mixture of iron gall and carbon ink. Initial words are often decorated in red and/or black, as are dragon-like and floral motifs. A few illuminated pages (*plate 1*, see p. 49) are painted in bright gouache, in blue, yellow, pink, brown and green colors.

The Seminary copy, which has 17 quires and 135 leaves, is probably missing the first quire. The pages measure 46.2 × 35.2 cm, generally with 26 lines per page. Over the course of 3 re-bindings, the pages have been cropped at the head and tail, and up to the prick marks on the outer margins.

The manuscript had six overlays adhered to the original pages. They were written in a greyish-toned ink and in a different hand from the original: three on goat parchment and three on calf.

Both volumes were bound in identical, eighteenth-century, tight-backed brown calf leather over pasteboard bindings. The New York manuscript had a detached upper board, a partly missing spine covering and unraveled sewing. More seriously, the inappropriate tight back with its glued up spine was causing cockling and distortion of the pages, potentially endangering the stability of the text and decoration. A decision was therefore made to re-bind the volume out of house, in a style compatible with the period of the manuscript; the disbinding and treatment of the pages were to be done in house.

The manuscript was, on the whole, reasonably well preserved. The first leaves were badly soiled, and the codex pages displayed general grime, accretions and stains. Cockling, creasing and flaking ink were apparent. Some of the holes and tears were in danger of being extended as the leaves were flexed and turned, and a few areas of corrosive ink which had eaten through the support caused concern. There was an abundance of white, brown and grey candle wax stains throughout the volume. Bloom on the text (especially on the large letters), which turned out to be the re-depositing of crumbled wax deposits, was a serious problem likely to get worse with the passage of time.

Tears and holes deemed unlikely to be extended were not touched. However, tears which had not been previously stitched or repaired, and holes caused by corroded ink, were repaired with gold-beater's skin. Goldbeater's skin [13], a transparent membrane prepared from the lining of a cow's stomach, is degreased with acetone and then rubbed with pumice powder or Fuller's earth. It can be toned, before mounting onto Japanese paper with a 3% solution of Klucel G (hydroxypropylcellulose) in ethanol for easier handling. An appropriate patch is cut out and adhered to the damaged area with parchment glue, usually from both sides. Later, the paper laminate can be removed, leaving a transparent repair. An infill of Japanese paper, such as Kitakata, can occasionally be sandwiched in between the two layers of goldbeater's skin.

The whitish bloom mentioned above, evident to the naked eye on most pages, especially on the oversized letters that frequently appear in the text, diminished the clarity of the scribe's clear black strokes on the cream parchment. When examined under a Nikon stereo microscope, it was clear that the deposits, sometimes in the form of a thin film, sometimes as tiny droplets, were caused by wax deposits. The obvious source was the cracking and powdering candle drippings evident on many pages. Although we had originally intended to leave the wax undisturbed, as evidence of the use and history of the *mahzor*, we decided to carefully remove the large drops of wax mechanically, with bamboo spatulas, to avoid future obfuscation of the text. Passing a Magic Rub eraser over the bloom of the large letters restored their black appearance, by consolidating the wax and changing the refractive index.

The six overlays previously mentioned had been stuck to the original leaves with animal glue adhesive [14]. Damp blotters were therefor applied to the overlays through a sandwich of hollytex and Gore-Tex, covered with Plexiglas and weights, and left to humidify and loosen the glue for two to five hours.

The overlays were then carefully peeled off to reveal the underlying text. As much of the thick layer of glue as possible was removed from both surfaces with Teflon spatulas and cotton swabs. The separated sheets were stretched and flattened between hollytex, dry blotters, Plexiglas and weights, leaving the conjoint leaf untouched.

The uncovering of the original text after hundreds of years seemed very dramatic to us, but it turned out that the hidden text contained only slight textual variations or

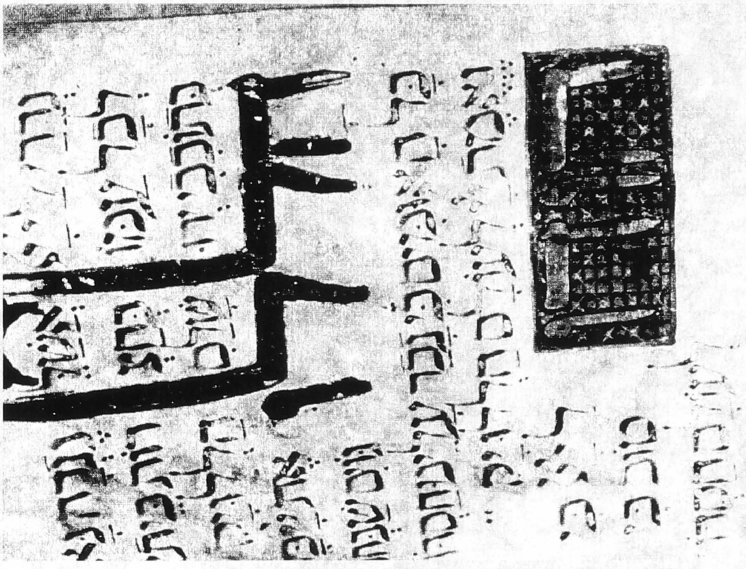


Fig. 4

Fig. 3

changes in the order of certain passages, possibly indicating that the codex had been physically removed to another town where the order of liturgy was slightly different.

The overlays were hinged onto their original positions by Deborah Evetts, thus enabling scholars to view both the original and emended texts. Evetts also executed a new binding, in a style appropriate to the period of the manuscript.

3. Graziano *Haggadah*

Probably written in Spain, at the beginning of the fourteenth century, the Graziano *Haggadah* takes its name from its eighteenth-century Italian owner, Rabbi Abraham Joseph Solomon Graziano of Modena. A description of the manuscript and its peregrinations prior to its acquisition by the Jewish Theological Seminary from the collection of Elkan Nathan Adler has appeared in a study by Evelyn M. Cohen [15]. The text, read at the festive *seder* table on Passover, recounts the biblical epic of the exodus of the Jewish people from Egypt. This particular manuscript, with initial letters, decorations (*plate 2*, see p. 52) and illustrations richly illuminated in gouache, silver and thickly-cushioned gold leaf on gesso, is written in black and brown ink in Sephardic square script. It consists of 35 parchment leaves, 25.0 × 19.0 cm, and had been re-bound in a brown leather binding. The *Haggadah* is heavily soiled and stained with wine, which possibly occurred during the *seder* when it is mandatory to drink four cups of wine.

The immediate and most serious problem was the extensive flaking and loss of both ink and pigment, dramatically visible during examination under a stereo microscope (*fig. 3*). Loose flakes of paint moved when barely touched by a soft sable brush or sharpened bamboo stick. The instructions of the curators were that both the pigments and ink be consolidated without disbinding the volume.

Parchment scraps were cut into small 1 cm squares and soaked overnight in distilled water. They were then covered with fresh water and simmered over low heat in a double boiler for about 5 hours, with the scum being constantly skimmed off. The resulting solution was filtered through several layers of cheesecloth, and poured into ice cube trays to set.

A book cradle was constructed out of foam core in order to support the *Haggadah* during consolidation, and a strip of curtain weights [16] was lightly draped over the open leaves to keep them in place and prevent them from disturbing the area being treated [17]. Parchment size was diluted until barely tacky when tested between the thumb and middle finger. A few drops of ethanol were added to break the surface tension. This solution was kept at a warm temperature in a small beaker that was placed in a small pyrex dish half filled with water placed on a small hot plate (of the sort usually used for keeping a cup of coffee warm).

The consolidation was carried out under a Nikon stereo microscope, at a magnification of between 10 to 30 times. A very fine, long-haired sable brush (size 00) introduced a small amount of ethanol into the area to be consolidated; this was followed by another brush loaded with warm parchment size. The brush was applied under the loose flakes of pigment and around the perimeter of the losses, and the relaxed and loose paint was drawn down to the parchment support by capillary action. Sometimes this procedure had to be repeated a few times. On the occasions when the flakes of paint did not return to the plane, they were coaxed into position with the help of a beveled minibamboo stick or Teflon spatula, through a strip of siliconized mylar, once they were almost dry. No size was brushed or sprayed on the surface of the manuscript. There was no visible change of color to the treated areas, neither was any surface sheen introduced (*fig. 4*). The success of the consolidation was monitored under the stereo microscope by gently running a pointed bamboo stick or porcupine quill over the treated area.

This method of consolidation is extremely time-consuming and requires a fair degree of skill and patience. The fact that parchment size is compatible with the original fabrication of the manuscript was a significant factor in our choice of consolidant, and the results of this painstaking process of conservation appear quite satisfactory.

Notes

1. Jay Rovner, "The computerized *genizah* cataloguing project of the Jewish Theological Seminary of America", *Shofar*, 8/4 (Summer, 1990).

2. A starch test with 0.13 g iodine in a solution of 2.6 g potassium iodide in 5 ml water, diluted to 100 ml before use, gave a positive result, indicated by a deep blue color. See "Spot tests in the American Institute for Conservation", *Paper Conservation Catalogue*, 7th edn. (October, 1990), p. 26.

3. Membrane of expanded polytetrafluoroethylene laminated onto polyester felt. W. L. Gore Associates Inc., Elkton, Md. 21921.

4. Purchased from TALAS. 568 Broadway, New York, N. Y. 10012. See *Conservation Catalogue*, op. cit., 8th edition (1992), pp. 22—3.

5. 400 centipoises. Sigma Chemical Company, P. O. B. 14508, St. Louis, Mo. 63178.

6. Sigma a-Amylase, Type IIA, Catalogue No. A-6380. Sigma Chemical Company.

7. T. Petukhova and S. D. Bonadies, "Sturgeon glue for painting consolidation in Russia", *Journal of the American Institute of Conservation*, 32/1 (1993), pp. 23—31, and Sarah Foskett, "An investigation into the properties of isinglass", *SSCR Journal*, 5/4 (November, 1994), pp. 11—4.

8. Polyester plastic film made by Du Pont.

9. Spun-bonded polyester, purchased from TALAS.

10. Purchased from Falkiner Fine Papers, 76 Southampton Row, London WC1 4AR, England.

11. The Pierpont Morgan Library, New York.

12. Evelyn M. Cohen and Emile G. L. Schrijver, "The Esslingen *Mahzor*: a description of the 'New Amsterdam' and 'Old Amsterdam' volumes", *Studia Rosenthaliana*, 25/1 (1991).

13. Purchased from Z. H. de Groot, Heemraadssingel 255a, 3023 Rotterdam.

14. A positive reaction to a Biuret test for the presence of protein (2% copper sulphate followed by 5% sodium hydroxide) gave a blue color.

15. Evelyn M. Cohen, "The Graziano *Haggadah*", *Outlook*, 64/3 (1994), pp. 16—8.

16. Purchased in the notions department of John Lewis, Oxford Street, London W1A1EX.

17. The techniques described were learned from Abigail Quandt, during an internship by Nellie Stavisky at the Walters Art Gallery, Baltimore, during the winter of 1995. Any errors of description are the responsibility of the authors alone.

Illustrations

Plate 1. Illuminated page of the Esslingen *Mahzor*, MS 9344, fol. 2r (see p. 49).
Plate 2. Graziano *Haggadah*, MS 9300, fol. 23v, decoration (see p. 52).

Fig. 1. Letter of Moses Maimonides to Rabbi Pinchas ha'Dayan in Alexandria, describing his extremely arduous daily schedule, MS 8254.14 (before treatment).
Fig. 2. The same letter (the process of inlaying; transmitted light).
Fig. 3. Microphotograph of flaking ink, Graziano *Haggadah*.
Fig. 4. Graziano *Haggadah*, fol. 32r (after consolidation).

Plate I

The table clearly shows that it is only the most common words that appear in a 100,000 words text with frequency big enough to draw some conclusions on their behaviour [16]. It is to be expected that less common verbs will need much bigger corpus to provide enough data on their use in the language.

These types of difficulties more or less determine the shape of a corpus of Arabic. It is obvious that for more sophisticated analysis, the corpus should be tagged, and the minimum requirements for the tags types are: (i) tagging morphological boundaries; (ii) part-of-speech tags; and (iii) providing the root information. The size of the corpus has to be relatively big, as showed the analysis of some characteristics of a 100,000 words text, which obviously provided enough information only on the most common words. The example of the Brown corpus of English (1 million words) shows that even such a size is not big enough for a proper analysis of a language, and in case of Arabic as a flectional language it is clear that the frequencies of especially verbs would be much less. It is quite probable that, e.g., for a lexical studies, even a corpus consisting of 10 million words might not be big enough.

This lead us to the decision to start work on a corpus of Arabic [17], aimed at modern standard Arabic, especially from the last 30 years. The projected size of the corpus is now 30 million words, and we assume that this size might be big enough even for lexical studies. The basic characteristics of the corpus would be: a balanced corpus with tags for morphological boundaries, part-of-speech, and root.

As the corpus is projected as a balanced one, we will try to cover as many varieties of Arabic as possible, i.e. we will gather texts from all major regions of Arabic, i.e. the Arabic Maghreb, Mashreq, and the Gulf area. It will cover both texts from periodicals (newspapers, magazines) and books, and will try to find a balance between various language styles.

Below, there is one of possible shapes of the corpus, certainly not free of problems and points that have to be further discussed.

Table 3

Number of the token	Token	Morphological boundaries	PoS tag [18]	Root
0001	وكان	و-كان	VPBe	كون
0002	الخلفاء	ال-خلفاء	NNP	خلف
0003	من	من	Prep	—
0004	الجهة	ال-جهة	NPS	وجه
0005	الاخرى	ال-اخرى	NAs	ءخر
0006	يعرفون	يعرفون	VIP3m	عرف
0007	حاجة	حاجة	NNS	حوج
0008	لأمراء	ال-أمراء	NNP	ءمر
0009	المسلمين	ال-مسلمين	NNP	سلم
..١٠	الى	الى	Prep	—
0011	رضاهم	رضا-هم	NNsP	رضو

Notes

1. A constantly growing commercial project of a monitoring corpus of English. Available at the University of Birmingham. A number of words in the corpus announced in summer 1996 was 320 million.

2. A project directed by the Oxford University Press, a balanced 100 million words corpus.

3. I.e. only the control characters are eliminated, only headlines and paragraphs are possibly marked.

4. The CALLHOME Egyptian Arabic corpus of telephone speech, available from the Linguistic Data Consortium, University of Philadelphia, consists of 120 unscripted telephone conversations between native speakers of Egyptian Colloquial Arabic. For more details, cf. the LDC Home page (http://www.ldc.upenn.edu).

5. This corpus has been developed by the Sakhr Company (Egypt, Saudi Arabia, (http://www.sakhr.com). According to my knowledge, it is available only internally for the company.

6. E.g., the 3rd version of Sakhr's Automatic Reader offers acceptable results even without the necessity of training the fonts. Besides, there are products offered by Caere (Arabic OmniPage) and TexPert for Macintosh. In the reviews that appeared in the electronic discussion lists (especially ITISALAT), the Sakhr's product seems to be superior to the other ones. According to my own experience, with quality printouts the success rate can reach 99%, requiring only very little postprocessing.

7. The last requirement is not really serious, since the character sequence on both DOS/Windows and Macintosh platforms more or less retain the character order of the Arabic alphabet.

8. The completely non-vocalized text in the extent of 1,000 graphemes resulted in our analysis in 1,584 graphemes of its fully vocalized counterpart, i.e. with the representation of all the short vowels, endings, and geminated consonants.

9. This might not be that serious for a linguist, but it is impractical in two aspects. First, the acquisition of new data would be very laborious, and secondly, any practical applications might fail to analyse real Arabic texts.

10. Kenneth R. Beesley, "Arabic finite-state morphological analysis and generation". Paper read at COLING-96, Copenhagen, August 1996, 6 pp.

11. The ambiguous cases can be quite numerous, for example, in Hans Wehr's dictionary, the roots beginning with *bj* are 8 and of them, 6 can be interpreted as consisting of the preposition *bi-* and a biradical root.

12. Hans Wehr, *A Dictionary of Modern Written Arabic*. An enlarged and improved version of Hans Wehr's Arabisches Woerterbuch für die Schriftsprache der Gegenwart, English translation by J. M. Cowan (Wiesbaden 1961—1994).

13. Beesley, "Arabic finite-state morphological analysis and generation".

14. The "token" here is understood as any string between two spaces. This certainly means that there are strings that contain more than one word, i.e. there are strings that consist of prefixes (prepositions, particles, etc.), word and suffixes (suffixed pronouns), as it has been described here above. Another fact worth of attention is that these tokens do not distinguish between various types of parts of speech, i.e. one token can represent both verbs and nouns. This has also been mentioned here above.

15. This number is a number of various verb forms appearing in the set. There are certainly strings that can be interpreted as both verbs or nouns, but since they can be interpreted as both, it can be assumed that these strings, at least to some extent, represent also verbs.

16. It is obvious that the types of verbs here correspond very strongly with the type of the text used for the collection of data. Most of the verbs are typical for a political news type of text.

17. From 1997, this project is supported by the Grant Agency of the Czech Republic, under the name *Thesaurus Linguae Arabicae*.

18. The tags used here are only provisional, there are still problems to be discussed. E.g., there is little difference between names and adjectives in Arabic, quite often a word can serve both as a noun or an adjective. Another problem is the representation of affixed words, and there are many other issues that will need a careful consideration.

Plate 2