William F. Friedman's Transcription of the Voynich Manuscript

Jim Reeds

AT&T Bell Laboratories

Murray Hill, New Jersey 07974

ABSTRACT

The Voynich Manuscript is a late medieval or early modern book written in an unknown cipher alphabet. It has resisted the efforts of several of the century's best cryptanalysts to break its cipher. One of them, William F. Friedman, prepared a machine readable transcription of this book half a century ago; this transcription has recently been unearthed from the archives and placed on line.

Introduction

For the better part of this century scholars have been puzzled by the mysterious, still-unread, Voynich cipher manuscript.* Notable among these was the eminent American cryptologist William F. Friedman, who organized an effort to transcribe and decipher the VMS at the end of World War II.

One of the difficulties facing anyone attempting to read the Voynich manuscript is the tedium of preparing a transcription into conventional alphabetic or numerical symbols, in order to make possible frequency counts and concordances. Such a transcription requires study of often unclear photocopies, and, if carried out with conscientious regard for accuracy, is very time

*Hereafter abbreviated ''VMS.''

7 September 1994

consuming. My experience suggests that after transcribing about ten pages the would-be VMS reader loses interest in the task, starts worrying about his eyesight and stops work, leaving more than two hundred pages untranscribed.

It was with great pleasure, then, that I discovered in the William F. Friedman Collection† of the George C. Marshall Foundation a transcription made half a century ago of almost all of the Voynich manuscript. This transcription, item 1609 in the Friedman Collection, is a printout of an IBM card ‘‘edition’’ of the VMS. It is the product of what M. E. D’Imperio [D1] calls the ‘‘First Voynich Manuscript Study Group’’ of 1944-46 (here abbreviated FSG), which was an unofficial after-hours club of U. S. Army cryptanalysts at the end of World War II. As far as I know, it is the only complete transcription of the whole VMS into conventional symbols, yet its existence has not been mentioned in print until now; D’Imperio [D1] seems unaware of its existence. It may well be the first example of a machine readable edition of a text prepared for scholarly purposes.

This paper describes item 1609 and places it in the context of the activities of the 1944-46 study group and in the context of Mr. and Mrs. Friedman’s longstanding interest in the Voynich manuscript, as revealed in the holdings of the Friedman Collection.

The present paper should be regarded as a contribution to the historiography — but not to the solution — of the Voynich manuscript. In a later paper I hope to present a statistical analysis of the VMS text itself, based on Friedman’s transcription.

The Voynich Manuscript

The following facts about the VMS are repeated in almost everything written on the subject. The VMS is a book of about 104 vellum leaves, sized about 16 cm by 23 cm (6 by 9 inches), written in an unknown (but apparently alphabetical) script. It is profusely illustrated with plant drawings, zodiacal diagrams, what have been called ‘‘cosmological’’ diagrams, and diagrams

†No general survey of the contents of this collection has been published. The items mentioned in this paper account for perhaps three quarters of the Voynich material in the collection; at a later time I hope to publish a systematic survey of the Voynich holdings.

with unclothed women romping on what seem to be water slides. The style of writing and of drawing seem to place its production in the late 1400s or early 1500s, possibly in central Europe.

The VMS was found in Italy in 1912 by the American rare book dealer Wilfrid Voynich [V, 1921]. A letter found with the manuscript, a faded signature on the first page, and a variety of collateral or circumstantial evidence definitely place the book at the Prague court of Rudolph II, Holy Roman Emperor from 1576 to 1611, and probably place it in the possession of the English mathematician and astrologer, John Dee, who was at Rudolph's court at various times between 1584 and 1588.

Ownership passed through Voynich's heirs to H. P. Krauss, another rare book dealer, who gave it to Yale University, where it is now MS 408 in the Beinecke Rare Book Library.

Since its discovery in modern times the problem of reading this book (or indeed, of making any sense out of it at all) has been a tantalizing puzzle to many scholars. About ten solutions have been offered in print, of varying degrees of implausibility. The claimed authors and topics include: God, Roger Bacon, Anthony Askham, Cathars, Khazars, spiral nebulae, contraceptives, suicide, capsicum, sunflowers and other botanical novelties from the New World. William R. Newbold's 1928 book *The Cipher of Roger Bacon* [N] contains the oldest and most notorious such fallacious solution, which was refuted by J. M. Manly [M]. There is a largely repetitious secondary literature on the VMS, a sampling of which is listed in the bibliography.

There are two serious books. Brumbaugh [Br] offers — with unconvincing evidence — the attractive theory that the book has no meaning, but was concocted as a hoax: a neo-Platonist rarity to sell to a wealthy gull, most likely the Emperor. In contrast, D'Imperio's book [D1] — an encyclopedic survey of everything known or conjectured (up to 1978) about the VMS — offers no solution of its own. Because D'Imperio was able to interview many of Friedman's Voynich collaborators (especially John Tiltman and Prescott Currier), her book is most useful as a guide to the Friedman collection; the preface by John Tiltman makes clear a form of ''apostolic succession'': D'Imperio succeeds Tiltman, who succeeded Friedman as ''unofficial coordinator of the work of some of the people who have been working on the problem.''

A privately printed pamphlet seminar proceedings by D'Imperio [D2] contains the best

statistical data about the VMS. The most accessible general account is perhaps in Kahn [K], pp.863-872. Clear reproductions of a few VMS pages may be found in [K], [N], [T], and [Z], but not in [Br] or [D1].

W. F. Friedman and the VMS

Elizebeth S. and William F. Friedman's longstanding interest in the Voynich manuscript is well reflected in the holdings of the Friedman collection in the Marshall Library in Lexington, Virginia.

They became interested in the VMS as soon as Newbold began publicizing his theories about the VMS in the 1920s, and started an extensive correspondence on the subject with their friend John M. Manly, the University of Chicago Chaucer expert who, like Friedman, had served as a cryptanalyst in World War I. Friedman obtained photostats from Mrs. Voynich, and through her, started a correspondence with Father Theodore C. Petersen which lasted until the latter's death in 1966. Petersen's hand-made tracing onto onion-skin paper is now item 1620 in the collection, and is amply described on page 41 of D'Imperio's book. After making the copy, Petersen prepared elaborate indexes and frequency counts (both into notebooks and onto index cards — all now in the Friedman collection — but in the process scribbled up his copy with underlinings, colored pencil marks, and so on, to the extent that photocopies of his copy are often hard to make out.

Friedman's extensive VMS photostat collection is based in part on Petersen's, but it is now impossible to give a precise acquisition date or provenance for each photostat. Some were obtained from Mrs. Voynich, some were copied from Petersen's copy of photostats, some were copied from published images of VMS pages.

A bound set of photostats, item 1600 in the collection, was either made for Father Petersen in 1931, or is a copy of that set. (See D'Imperio, pp. 39 and 41 for slightly conflicting accounts; Tiltman [T, p.44] asserts that Friedman's photostats were copied from Petersen's, and Tiltman's from Friedman's.) It consists of 125 sheets of photographs, printed on large sheets of stiff photographic paper of varying sizes. One or two pages of the VMS is shown on each sheet, each page

marked with a small penciled, circled, page number. On most of the photostats a patch of black is visible, on which a folio number has been written in white ink*, in yet a different handwriting; possibly Voynich's. The first photostat in the bundle is a negative reproduction of Newbold's Plate XXI, with Friedman's handwritten note ''Note: This is a page which I have added as a protection to f1 of the original. Taken from the Newbold-Kent book on the Voynich MSS. W.F.F.'' The next photostat shows the first page of the VMS, has a circled ''1'' and the white ink label reads ''Fol. 1 recto,'' and so on. The photostats D'Imperio herself used, described on page 40 of her book, defaced by

> copious and obtrusive remains of at least one previous computer processing project, including circled words and paragraphs, lines marking off parts of the text, and legends such as ''start here,'' ''omit punch,'' and ''punch just this''

were apparently derived from the set in the Friedman collection, which are not as severely defaced, but which do have such legends as ''omit punch (entire page)'' on photostat page 112 (showing f.57v), a ''punch just this'' on photostat page 119 (f.67r), etc. She might have been using Tiltman's set [T, p.44].

Towards the end of World War II Friedman's Voynich investigations took a more serious turn, with the formation of what D'Imperio called the ''First Voynich Manuscript Study Group,'' (FSG) an after-hours informal club of Army cryptanalysts. This club was apparently active between 1944 and 1946; the results of its activity are described below. In 1946 Friedman obtained a report (in folder 1614) on the VMS's handwriting from his colleague Albert Howard Carter (later professor of literature at Eckerd College).

In 1951 Friedman was able to interest John Tiltman, a British cryptanalyst he had befriended during the war, in the VMS. It seems that Tiltman became Friedman's closest confidant, at least in Voynichological matters. In the early 1950s Friedman developed his theory that

*Many of the VMS leaves are foldouts, and the existing folio numbers do not all lie on the recto sides of the leaves. Hence it is not always obvious what to call any given page of the VMS; the foliation used in [N] and [Br] is often haphazard. The ''white ink'' folio numbers shown on item 1600 provide the most systematic and unambiguous method of naming the pages of the VMS that I know of; this is the foliation used by D'Imperio [D1]. The pencil page numbers were used in all computer and punch card projects by Friedman and associates. An appendix shows how the two systems are related.

the VMS was written in an artificial language, similar to those of Cave Beck and John Wilkins [Z], [T]; this was published in the form of an anagram announcement in a footnote to a 1959 *Philological Quarterly* article ‘‘Acrostics, Anagrams, and Chaucer.’’ Folder 1614 contains an exchange of letters in 1954 with Erwin Panofsky about the VMS.

For much of the 1950s the Friedmans were occupied with the studies leading up to the publication in 1957 of their *The Shakespearean Ciphers Examined.*

In 1962 Friedman was able to assemble another after-hours group, the ‘‘Second Voynich Manuscript Study Group’’ (SSG), which sought to enter the text of the VMS into an RCA computer. The Friedman collection possesses several SSG items: item 1609.4, which includes a file of memos (of a highly bureaucratic nature, about the protocol for use of RCA equipment) and alphabet sheets; and a 63-page printout of a computer transcription into 46,424 computer characters of VMS pages 120 through 175 (which are f.67r1 through f.87r); and item 1609.3, a massive 692 page computer printout of a cross reference or ‘‘KWIC’’ tabulation of the transcription in item 1609.4. At roughly the same time Elizebeth S. Friedman wrote a survey article [F] about the VMS, published in the *Washington Post*, and William, possibly anticipating progress from the SSG, planned a weightier article for publication in *Isis* or *Speculum.* Unfortunately, the SSG effort terminated before it had much to show for its efforts, and the weightier paper was never written.

At about this time Friedman’s health began to deteriorate and he did no further major work on the VMS in the final decade of his life.

The First Study Group

There seem to be no published first-hand accounts of the activities of the First Study Group (hereafter abbreviated FSG). More-or-less equivalent secondary accounts can be found in Kahn [K], Zimansky [Z], and Clark [C]; a slightly more detailed account in D’Imperio [D1] (who had access to a partial set of minutes of the FSG’s meetings): At the end of the war, the Army cryptanalysts headed by Friedman found themselves without any pressing tasks. Many were simply awaiting demobilization and return to their universities and civilian practices. Friedman took

advantage of their momentarily free time and talent by organizing an effort to work on the Voynich problem. The group studied the available scholarly material, discussed hypotheses, transcribed the VMS onto IBM cards, and disbanded.

Unknown are: the location of the archives of the FSG, the membership of the FSG, the current location of their IBM cards. It is possible that some of this material is in still-classified archives of the NSA. It is known that Frank Lewis (personal communication with James Gillogly, 1993) and Martin Joos (personal communication, ca. 1968) were in the right place at the right time to have been part of the FSG but Lewis was not attracted to the Voynich problem and Joos thought Friedman's approach was misguided, so neither participated.

The Friedman Collection's FSG materials are confined to printouts of IBM cards, alphabet sheets for transcribers, and worksheets; they contain no narrative or administrative material of the sort cited by D'Imperio. (Some of the worksheets, however, bear signatures, most of which must belong to FSG members.)

FSG Alphabet Sheets

Before they could start their main work of transcribing, the FSG had to pick a transcription alphabet. This involved two choices: they had to settle on what they thought the Voynich character set was, and they had to establish conventional letter or number equivalents for each Voynich character.

Of these two choices the first is the more critical, for it determines the level of detail, and the kinds of detail, with which the VMS will be transcribed. The kinds of mistakes that the wrong choice leads to can be imagined by supposing a future race of beings trying to decode our writing system. If they mistakenly assume that ''m'' and ''n'' are the same letter (because they don't believe the exact number of humps could be important) or that ''h'' and ''n'' are the same letter (because they differ only in length of a single stroke), or that ''n'' and ''u'' are the same (because they are rotated versions of each other) their analysis will be made harder. On the other hand, if they think that ''m'' and ''*m*'' are genuinely different letters, or that ''A'' is fundamentally different from ''a,'' their analysis might become bogged down with irrelevant minutiae.

The choices the FSG made about its transcription alphabet are recorded on a sheet of paper found in file 1609.2, which is partly typed (or mimeographed) and partly filled in by hand. The typed part of the sheet consists of a heading, ‘‘TENTATIVE LIST OF CHARACTERS,’’ and two columns of blanks, numbered 1 through 14 on the left and 15 through 28 on the right. The handwritten marks on the document are most interesting. Some of the numbers are crossed off, some are renumbered. To the left of most numbers stands a handwritten letter, to the right stands a hand-drawn Voynich character and another letter. The heading is annotated ‘‘As agreed at Meeting on 9 June 44’’, and the document is signed ‘‘Transcribed by W.F.F. 13 June 44.’’ The overall appearance of the document is somewhat like the following:

TENTATIVE LIST OF CHARACTERS

A	1.		P	O	15.		4
B	2.		F	P	16.		E
C	3.		H	Q	17.		C
D	4.		D				
E	5.		G	R	18.		T
F	6.		A	S	19.		S
G	7.		R	T	20.		I
H	8.		K	U	23.		HZ
I	9.		2	V	21.		PZ
J	10.		O	W	24.		DZ
K	11.		L	X	22.		FZ
L	12.		N	Y	25.		V
M	13.		M	Z	26.		Y
N	14.		8				
		27.	, space				
		28.	,, par				
		29.	ø illegible character				

Thus, on 9 June 1944 the FSG decided that the VMS character set consisted of 26 characters, had decided to record word spaces, breaks between paragraphs, and illegible characters, but decided not to record ends of lines. Further, they picked three assignments of conventional values to the Voynich letters. We may term these, reading left to right, the ‘‘alphabetical,’’ the ‘‘numerical,’’ and the ‘‘mixed’’ or ‘‘mnemonic’’ values.

Several mimeographed copies of another alphabet sheet are found in file 1609.1. They look approximately like this:

VOYNICH MANUSCRIPT

ALPHABET FOR TRANSCRIPTION

1.		P	16.		E
2.		F	17.		C
3.		H	18.		T
4.		D	19.		S
5.		G	20.		I
6.		A	21.		PZ
7.		R	22.		FZ
8.		K	23.		HZ
9.		2	24.		DZ
10.		O	25.		V
11.		L	26.		Y
12.		N	27.	,	space
13.		M	28.	,,	paragraph
14.		8	29.	ø	illegible charac.
15.		4			

and bear the mimeographed signature of Mark Rhoads, an army cryptanalyst. One is dated (in Friedman's handwriting) 6 Jan 1946, but it could have been typed any time after 13 June 1944, since it is simply a tidy restatement of the previous alphabet sheet, with the ''alphabetic'' values dropped.

Description of Item 1609

The printout consists of a bundle, fraying at the edges, of 130 sheets of printout paper, originally part of a continuous web, but since burst into separate sheets. The sheets are 11 inches wide and 17 inches long. On the lower left edge of each sheet the notations ''TABULATING PAPER

- NO. 1'' and ''SIS - SC - 861'' are printed in very small letters.

Each sheet except the last has 46 lines of printing; the last has 40, making 5974 lines of printing in all. It appears to be a listing of a deck of IBM cards. The printing is spaced three lines to the inch. Each line of printing is about 6.5 inches long, with an ample margin on the left side of the sheet. Each line consists of a group of 5 digits, a space, a group of 3 digits, two spaces, and then as many as 30 letters, digits, and commas in the mnemonic transcription alphabet. With the exception of a few sheets on which the printer ribbon seems worn out, the printing is quite clear. One distinctive feature is the form of the printed zero, which looks more like Norwegian ø than an ordinary 0.

The first group of digits names the VMS page, the second group is a serial number for IBM card within page, and the 30 letters, digits, and commas are the actual transcription data.

There are, in addition, two sheets of handwritten transcription data, clipped to sheets 20 and 22.

Many sheets have extensive handwritten marks on them. These may be roughly classified as marks of ownership, of use, and of correction.

The ''marks of ownership'' are mainly on the first sheet: At the top of the sheet, in pencil cursive, not Friedman's, is the misspelled label ''Voyanich Man.'' and (in another hand) ''Rhoades H. Q. 215,'' doubtless referring to Captain Mark Rhoads (USA, Ret.), Friedman's long-time associate [C, 1977, see pp 87, 88] and apparently, at the end of the war, Friedman's chief assistant. In the upper right hand corner of the the first sheet is a rubber-stamped date, ''Jan 07 1946,'' which is consistent with the 1944-1946 date span usually cited for the First Study Group. Just above the first line of printing, in pencil block letters is ''Tentative IBM Transcript.'' Immediately to the left of the first printed line, in pencil cursive (possibly Friedman's) ''27 Mar. '49.'' On the bottom of the first sheet, a large handwritten ''1609.'' On sheet 3, another ''27 Mar '49.'' Starting on the 87th sheet of printout are comments in the left margin, in Friedman's hand, written in green ink, like ''Here begins the 'Rx's' or 'recipes.' (Folio 103 recto),'' and at the top of each sheet from there on to the end, a catch phrase like ''F 103 R Cont'd.'' On sheet 96, ''Begin F.105 Verso. Note: Tiltman's transcript begins here.'' On sheet 119, ''Tiltman's

transcript stops here, at end of ¶ 12.'' And at the bottom of sheet 130, the last: ''F. 116 V — last page. The last page, with only three lines of writing, is alleged by Newbold to contain 'the key.'''

The ''marks of correction'' are found on blocks of sheets throughout the printout. Sheets 1 through 34, sheets 46 through 51, sheet 96, sheets 117 through 119 all bear hand corrections. These parts of the printout look like marked up author's proof sheets: letters are crossed out, over-written, inserted with carets. In many cases line breaks, and line and paragraph numbers have been indicated in pencil, with notes like ''¶ 6'' inserted into the body of the text.

And finally, the printout shows signs of use. Every character on the first 33 sheets and first nine lines of sheet 34 has been underlined in pencil, each separate letter or digit receiving its own separate stroke (except that both letters in the digraphs PZ, FZ, HZ, and DZ are underlined by the same stroke). These strokes could be the byproduct of preparing a frequency count or of proof-reading; I am inclined to believe the latter. On sheets 50 and 51 there is evidence of frequency count taking: underlined letter sequences, with a marginal note ''4ODC8G 15 times on p 153'' on sheet 50.

Folder 1613

One item in the Friedman collection, folder 1613, contains a miscellany of worksheets, apparently from the FSG. Most of these are on sheets of 1/4 inch graph paper, by a variety of writers, containing frequency tabulations of Voynich characters, of letters in medieval Latin texts, and the like. Many bear signatures, which are our only clues to the rank-and-file membership of the FSG. Some of these names are (as far as I can read them): Robert A. Caldwell, G. E. McCracken, Thomas A. Miller, William M. Seaman, Fried, and Francis M. Puckett.

One item in folder 1613 is especially interesting: another partial transcription of the VMS in a slightly different form from that in item 1609. It consists of twenty 8½ by 11 inch sheets of 1/4 inch graph paper, with a handwritten transcription of f.111v through f.114r. (''Pages'' 225 through 230.) Each sheet (turned sideways) has a block of transcription data, 40 squares wide by 20 squares high, one character per square, making 15,464 transcribed characters (the last sheet

falling short). The sheets are numbered serially, 1 through 20, and the corresponding VMS page is written inaccurately on each sheet. The first four sheets are labeled ‘‘Folio 112 Verso’’ when in fact they transcribe f.111v, the next three are correctly labeled ‘‘Folio 112 Recto’’, the next few incorrectly ‘‘Folio 113 Verso’’ and so on, systematically mislabeling all verso sides. The first sheet has the words ‘‘Francis M. Puckett, Lot #4’’ written on one edge and in blue pencil the words ‘‘To be verf,’’ the first two of which are crossed out.

These sheets are evidently IBM card coding sheets, that is, the written matter an IBM key punch operator looks at while keying in the data. The transcription data has been carefully marked out into 30 character blocks, with a heavy black stroke at the beginning of each block; each block has a 3 digit serial number ranging from 001 through 516.

These are not, however, coding sheets left over from the preparation of 1609. First, the card number series does not start over with each transcribed page, as they do in 1609 (which covers the same part of the VMS in 527 cards). Second, the transcription alphabet used is the ‘‘alphabetical’’ version seen on 1609.2, and not the ‘‘mnemonic’’ version used in 1609.

It is interesting to speculate about the reason for use of the variant transcription alphabet. It might be that this transcription was made before the ‘‘mnemonic’’ equivalents had been chosen, that is, before 9 June 1944. (The 1613 transcription is undated.) Alternatively, the 1613 transcription might have been carried out as an after-the-fact test of the transcription process, as a check on the accuracy of 1609. The variant transcription alphabet was chosen to deliberately shake the transcriber(s) out of any ruts of habit their transcription method might have fallen into, to force them to constantly refer to a non-standard and hence unfamiliar alphabet chart, in order to enforce greater accuracy on this ‘‘quality control’’ transcription.

To see how much 1613 differed from 1609, I instructed my computer to recast 1613’s transcription as much as possible into 1609’s terms. This is a preliminary brief report of the differences, given in 1609’s transcription alphabet. First, the recast version of 1613 takes 15580 characters while that portion of 1609 covering the same material takes 15682 characters. Of these, 14817 were the same in homologous passages. The 1613 version has 763 characters in places where the transcriptions differ, and the 1609 version has 865. This means they disagree by 5% or

6%. Most of the discrepancies involve single-letter mistakes: 208 confusions of M and N, 109 disagreements about the presence of word spaces, 71 confusions of A with O, and so on. There are about 100 one-of-a-kind single letter discrepancies. In addition, there seems to be a line of Voynich text recorded in 1609 missing from 1613, which accounts for much of the over-all length discrepancy.

Another partial FSG transcription

Also in the Friedman collection is an unnumbered item, closely related to item 1609. It consists of 26 sheets of printout paper of the same sort as seen in item 1609, protected with cardstock end sheets. Written on one of the protective end sheets are the words

pp. 79-113
Study worksheets for Voynich MS
Voynich Manuscript
Wilford CSB
X256

and on the first sheet in (a different hand) in large letters

Proof List — Page Nos, 79-113
Voynich Manuscript #2
Job # 1574.
Mrs Wilford Ext. 256

and a rubber-stamped date ‘‘SEP 24 1945’’ with the number ‘‘9473’’ written under that.

The sheets are ruled with a one inch space between rules; ‘‘TABULATING PAPER - No. 1’’ is printed in small letters on the left margins. The remaining 25 sheets have a transcription of pages 79 through 113 of the VMS (that is, f.41r through f.58r) in much the format that is used in item 1609: a five digit page number, padded with leading zeros, a line number (blank padded, taking up four spaces in all), a space, and 30 letters, digits, and commas of transcription data, also using the FSG mnemonic system, with two lines of printing per vertical inch. After each

Voynich page's transcription data is a single line showing the number of cards taken to transcribe that page, and four blank lines, giving a generous white gap between successive pages. At the very end is a line with the number ''738,'' which is puzzling, since 638 cards are listed in the printout.

A few of the sheets are lightly corrected by hand.

This is almost certainly a draft of part of item 1609. Oddly, it has a transcription of page 106 (f.54v), which is lacking from item 1609.

Comparison with Currier's Transcription

Some time in the 1970s Prescott Currier prepared a machine transcription of a large part of the VMS (pages 1 through 111 and 147 through 166, which is f.1r through f.57r and f.75r through f.84v), using a transcription system which recorded line ends. (This transcription is available by ''ftp'' computer network connections as file `/pub/voynich/voynich.orig` on the `rand.org` computer in Santa Monica, California.) Using this transcription, Currier made two very interesting discoveries, reported in [D2]. (1) The VMS language statistics depend on position in line: certain characters show a preference for the beginnings of lines, etc, and (2) there are two somewhat differing handwriting styles present in the book, with corresponding slight differences in letter and digraph frequency counts, as if the book were the product of two scribes with different handwritings and language usage statistics.

It would be desirable to make a close comparison of 1609 and Currier's transcription. A brief look at the differences yields the following: Currier's transcription, when recast into 1609 terms, is 85,124 characters long; the corresponding portion of 1609 is 85,357 characters long. They agree in 78,739 characters in homologous passages, they differ in 5,861 places in the text, where the Currier version has 6,385 characters and the item 1609 version has 6,618. Thus, the overall discrepancy rate is about 8%. Of the 5,861 places where the two transcriptions differ, about 2,940 only involve ''punctuation'' marks: word spaces, line and paragraph ends. In an attempt to find out which was the more accurate transcription, I randomly selected 120 of the 5,861 discrepancies, automatically skipping those solely concerned with line ends. Then I tried to

account for all 120 discrepancies by consulting photocopies. In 9 cases the problem was caused by unusual rare Voynich characters not found on the alphabet sheets. In 40 cases Currier's version seemed clearly correct. In 54 cases item 1609's version seemed clearly correct. In 13 cases both versions seemed wrong. In 4 cases the discrepancy seemed to hinge on interpretation of a line end or a very big word space caused by an intruding picture.

Conclusion

The transcription of 1609 is not very accurate*, and should probably be carefully checked and revised before it is used as the basis for statistical investigations. However, because it covers the whole VMS, it has some value as a ''base line'': if some other transcription of any part of the VMS is produced, it can be compared with 1609, and special attention can be paid to points of difference: wherever the two versions differ, one consults the photostats. This can be done right now with 1613 and with D'Imperio's transcriptions, which would give a fairly automatic way of checking about half of 1609. More precisely: 1609 contains 6030 lines, of which 2882 are covered by the D'Imperio transcription and 527 by the 1613 transcription, which leaves 2621 lines uncovered.

The omission of line ends is now known — because of Currier's findings [D2] — to be a bad mistake. (But editing them in should not be too hard.)

Except for the omission of line ends, the FSG transcription alphabet itself is not bad, and in one respect is superior to Currier's: it has a code for the Voynich character ⟙ which occurs often enough to deserve one of its own.

A curious side light was provided in a personal communication from Prescott Currier, who played a leading role in the Second Study Group of 1962: Friedman kept Currier in the dark about the FSG. Currier was not told what the FSG had accomplished nor did he see any transcribed text from the earlier effort. Is it possible that Friedman was somewhat ashamed of the

*This reflects no discredit on the FSG, given the difficulty of their task and the limited time available to work on it. In my experience proofreading a VMS transcription is harder than making a fresh transcription, and a 5% error rate is not unreasonable.

results of the FSG, whose transcription was not up to his usual standards of accuracy? A brief survey of Friedman's papers uncovered no written indication of such a feeling, and John Tiltman, with whom he might have confided, is no longer alive to tell us.

7 September 1994

Appendix: On-line version of 1609

I have prepared a computerized copy of item 1609 which incorporates all handwritten changes found on the printout. I have silently corrected a few mistakes in the page and line fields of the transcription. I have attempted to retain all the handwritten ‘‘marks of ownership’’ and ‘‘marks of correction’’ but not ‘‘marks of use’’ in the copy. In addition, occasional comments of my own have been added, marked in a way which makes it clear that they do not appear in the original.

In more detail: the transcription is a computer file, with three kinds of lines of text. Lines whose first character is a ‘‘sharp’’ sign, #, are my modern (1993-4) comments, which show the breaks between successive printout sheets, tell when a reading is doubtful, tell when an illegal transcription character indeed appears in the transcription data, and so on. Lines whose first character is a semicolon, `;`, are comments found on the original, usually in the margin of the printout sheet. All other lines are transcription data, either as printed in the original, or as indicated by handwritten corrections to the original. Within such a line, comments found in the original which can be localized to a particular place in the line are set off by a pair of matching curly brackets, `{}`.

The ‘‘punctuation marks’’ in the transcription data have been modified to follow handwritten corrections, as follows. The original FSG plan recorded word and line breaks with a single comma and page and paragraph breaks with a pair of commas. On all sheets in which corrections were made, however, the correctors inserted marks for line breaks, and usually also inserted line numbers and paragraph numbers, but with no uniform system of notation. Occasionally line breaks are represented by a single or double vertical stroke overwriting the printed comma in the original, and paragraph breaks are marked with a double or treble vertical stroke, overwriting the pair of printed commas in the original, but occasionally a ¶ sign is used. When present, line and paragraph numbers were indicated by superscript numerals added to the symbol used for line or paragraph break. I have rendered all added line-break symbols with a hyphen, `-`, and all added paragraph symbols with an equals sign, `=`. Whenever the corrector added a line or paragraph number I have put it in curly brackets. Whenever the corrector prefixed a paragraph number with

a paragraph sign, ¶, I used a dollar sign, $.

My method of making a modern computer transcription of item 1609 was as follows. The Marshall Library provided me with two xerographic copies of all of 1609, one of which I supplied to a volunteer, Jacques Guy. Working independently on separate continents, by different methods, we prepared two computer transcriptions of 1609. Then the two transcriptions were compared and all discrepancies resolved.

One of us (Guy), typed about two thirds of the text into his computer the ‘‘old fashioned’’ way.

The other (Reeds) tried to use modern optical character recognition (OCR) methods to scan his copy into the computer. In consultation with Henry Baird, an OCR expert, it became clear that the variations in inking of the original and in the quality of the photocopies, as well as the extensive pencil markings on some of the pages of the original, would make the output of an automatic OCR run unusably inaccurate, even if the OCR algorithm [Ba] were trained on the document itself. Instead, we devised a semi-automatic scheme, where the computer assigned a preliminary guess at the values of each of the printed characters on a page. Then the computer displayed images of all characters assigned value ‘‘A,’’ say, and the human operator could quickly spot and correct misassignments by using mouse clicks. As bogus As were reassigned, their images disappeared from the screen, leaving behind a ‘‘purer’’ field of As in which it was ever easier to spot misassignments. When all the putative As were indeed visibly As the computer then displayed all putative Bs, and the process repeated. At the end, the computer (which of course remembered where on the page each of the characters was printed) wrote a file showing the resulting reading of the given page of 1609. By using this technique, I did not see characters in context, which had both good and bad consequences. On the one hand, contextual clues for guessing at faintly printed letters were absent, so the error rate for such letters was elevated. On the other hand, errors which might have been introduced by paying less attention to a character than to its neighbors (analogous to unconscious correction of spelling mistakes in transcribing ordinary text, as well as transposition errors) were presumably avoided.

The results of the comparison of the two transcriptions are are as follows. In all, 113,366

characters (exclusive of card numbers, etc) were covered by both transcription methods. There were 136 differences between Guy's version and the version finally adapted and 78 differences between Reeds's version and the final version, not counting differences involving only the punctuation marks - and =. Thus, Guy and Reeds had transcription error rates of about .0012 and .0007, respectively. According to a naive (independent transcription errors) model, one might expect less than one lingering error in the portion transcribed by both methods.

The portion of Reeds's version *not* covered by Guy's (65,260 characters) was carefully proofread twice, first the old fashioned way, and then again while a computer-driven speech synthesizer read the transcription aloud. The second stage caught 44 errors, which means an error rate of about .0007; the naive error model again predicts less than one lingering error.

Even without accepting the sanguine predictions of the naive error model, it seems certain that Guy and I have introduced far fewer errors in making our modern computer copy of item 1609 than the FSG originally made in transcribing the VMS.

Appendix: Names of VMS pages

The page numbers given in the FSG transcription are the same as the circled numbers on the photostat set in the Friedman collection (item 1600). Here are the conventional (‘‘white ink’’) names for those pages, presented in tabular form. The sheet with pages 178 and 179 seems to be missing. It presumably shows more of the 9 medallion diagram of f.85/86.

-	0	1	2	3	4	5	6	7	8	9
0	-	f.1r	f.1v	f.2r	f.2v	f.3r	f.3v	f.4r	f.4v	f.5r
10	f.5v	f.6r	f.6v	f.7r	f.7v	f.8r	f.8v	f.9r	f.9v	f.10r
20	f.10v	f.11r	f.11v	f.13r	f.13v	f.14r	f.14v	f.15r	f.15v	f.16r
30	f.16v	f.17r	f.17v	f.18r	f.18v	f.19r	f.19v	f.20r	f.20v	f.21r
40	f.21v	f.22r	f.22v	f.23r	f.23v	f.24r	f.24v	f.25r	f.25v	f.26r
50	f.26v	f.27r	f.27v	f.28r	f.28v	f.29r	f.29v	f.30r	f.30v	f.31r
60	f.31v	f.32r	f.32v	f.33r	f.33v	f.34r	f.34v	f.35r	f.35v	f.36r
70	f.36v	f.37r	f.37v	f.38r	f.38v	f.39r	f.39v	f.40r	f.40v	f.41r
80	f.41v	f.42r	f.42v	f.43r	f.43v	f.44r	f.44v	f.45r	f.45v	f.46r
90	f.46v	f.47r	f.47v	f.48r	f.48v	f.49r	f.49v	f.50r	f.50v	f.51r
100	f.51v	f.52r	f.52v	f.53r	f.53v	f.54r	f.54v	f.55r	f.55v	f.56r
110	f.56v	f.57r	f.57v	f.58r	f.58v	f.65r	f.65v	f.66r	f.66v	f.67r1
120	f.67r1	f.67r2	f.67v2	f.67v1	f.67.r2	f.68r1	f.68r2	f.68r3	f.68v3	f.68v2
130	f.68v1	f.69r	f.69v	f.70r1	f.70r2	f.70v2	f.70v1	f.71r	f.71v	f.72r1
140	f.72r2	f.72r3	f.72v3	f.72v2	f.72v1	f.73r	f.73v	f.75r	f.75v	f.76r
150	f.76v	f.77r	f.77v	f.78r	f.78v	f.79r	f.79v	f.80r	f.80v	f.81r
160	f.81v	f.82r	f.82v	f.83r	f.83v	f.84r	f.84v	f.85/86r1	f.85/86v4	f.85/86v6
170	blank	f.85/86r2	f.85/86v5	f.85/86v3	f.85/86v3	f.87r	f.85/86v1,r3	f.85/86v2,r4	-	-
180	f.87v	f.88r	f.88v	f.89r1	f.89r2	f.89v2	f.89v1	f.90r1	f.90r2	f.90v2
190	f.90v1	f.93r	f.93v	f.94r	f.94v	f.95r1	f.95r2	f.95v2	f.95v1	f.96r
200	f.96v	f.99r	f.99v	f.100r	f.100v	f.101r2	f.101v2	f.101v1	f.102r1	f.102r2
210	f.102v2	f.102v1	f.103r	f.103v	f.104r	f.104v	f.105r	f.105v	f.106r	f.106v
220	f.107r	f.107v	f.108r	f.108v	f.111r	f.111v	f.112r	f.112v	f.113r	f.113v
230	f.114r	f.114v	f.115r	f.115v	f.116r	f.116v	-	-	-	-

Thus, page 143 is f.72v2.

Acknowledgements

I am, in the first place, most grateful to the George C. Marshall Foundation and Research Library, of Lexington, Virginia, for permitting us to examine, copy, transcribe, and quote material in its Friedman Collection, and to its staff, whose cheerfulness and intelligence makes this library a model of what a research library should be.

I am also grateful to Jacques Guy — a modern Herakles or Sisyphus — for undertaking the immensely tedious task of typing the greater part of the text of item 1609 into the computer.

Henry Baird was very generous with OCR support and advice.

I am also indebted to the following for personal communications: A. H. Carter, III, Mrs. M. Carter, Prescott Currier, M. E. D'Imperio, James Gillogly, the late Martin Joos, Susan Hockey, Louis Kruh, Albert C. Leighton, Frank Lewis, Margaret Zimansky.

References

Ba Baird, H. S. ''Anatomy of a Versatile Page Reader,'' *IEEE Proceedings, Special Issue on OCR.* July, 1992.

Br Brumbaugh, Robert S. *The World's Most Mysterious Manuscript.* Carbondale: Southern Illinois University Press, 1978.

C Clark, Ronald W. *The Man Who Broke Purple: The life of the world's greatest cryptologist, Colonel William F. Friedman.* London: Weidenfeld and Nicholson, 1977.

D1 D'Imperio, M. E. *The Voynich Manuscript: An Elegant Enigma.* Fort George G. Meade, Maryland: National Security Agency, 1978. (Reprinted by Aegean Park Press, Laguna Hills, California, c. 1980.)

D2 D'Imperio, M. E., editor. ''New Research on the Voynich Manuscript: Proceedings of a Seminar, 30 November 1976'' Privately printed pamphlet, Washington, D.C., 1976.

F E. S. Friedman. '' 'The Most Mysterious Manuscript' Still an Enigma.'' *Washington Post,* Aug. 5, 1962, sec E, pp. 1, 5.

K Kahn, David. *The Codebreakers.* New York: Macmillan, 1967.

M Manly, John Matthews, ‘‘Roger Bacon and the Voynich MS,’’ *Speculum* VI(1931), pp. 345-391.

N Newbold, William Romaine. *The Cipher of Roger Bacon.* Edited with forward and notes by Roland Grubb Kent. Philadelphia: University of Pennsylvania Press, 1928.

T Tiltman, John. ‘‘The Voynich Manuscript ‘‘The Most Mysterious Manuscript in the World’’.’’ *NSA Technical Journal.* XII, 3 (July 1967), pp. 41-85.

V Voynich, Wilfrid M. ‘‘A preliminary sketch of the history of the Roger Bacon Cipher Manuscript.’’ *Transactions of the College of Physicians of Philadelphia*, 43(1921), pp. 415-430.

Z Zimansky, Curt A. ‘‘William F. Friedman and the Voynich Manuscript,’’ *Philological Quarterly*. XLIX(1970), pp. 433-443.