

Jason B. Morningstar. Gallows Variants as Null Characters in the Voynich Manuscript. A Master's paper for the M.S. in I.S. degree. April, 2001. XX pages. Advisor: Gregory B. Newby

This study intended to determine how the elimination of gallows variants from the transcription set change the results of statistical queries on the Voynich manuscript. It was hypothesized that the gallows variants in the Voynich manuscript alphabet are null characters, and that removing them would not have a statistically relevant impact on correlational power curves. Voynich-based text samples were created that manipulated and removed gallows variants in different ways. These were analyzed and compared to the original text, looking for similarity and divergence. The actual analysis was a straightforward application of Spearman's rank correlation coefficient to nine separate data samples, along with the source text and two natural language control files written in vulgate Latin and Arabic, respectively. The study demonstrated that the removal of gallows variants effected the statistical measures in ways inconsistent with null characters.

HEADINGS

Correlation Analysis

Ranking

Term Frequency

Text Analysis

Voynich manuscript

GALLOWS VARIANTS AS NULL CHARACTERS IN THE VOYNICH MANUSCRIPT

by
Jason B. Morningstar

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Information Science.

Chapel Hill, North Carolina

April, 2001

Approved by:

Advisor

AUTHOR NOTE

I conducted this research as a Master of Science in Information Science student at the School of Information and Library Science, University of North Carolina - Chapel Hill.

I'd like to thank Jack Dagg, Robert Losee, Kevin Morgan, Greg Newby, Jim Reeds, Diane Sonnenwald, Lucia Zonn, the IRSS stat lab staff, the Voynich mailing list participants, and Mark Koyanagi, who bet me dinner I wouldn't finish on time.

Correspondence concerning this study should be addressed to Jason Morningstar, 4915 South Mill Road, Dryden, MI 48428. Electronic mail may be sent via Internet to mornj@ils.unc.edu. I can also be reached via the Voynich electronic mailing list at voynich@rand.org.

CONTENTS

LIST OF FIGURES	iii
INTRODUCTION	1
LITERATURE REVIEW	3
METHODOLOGY	11
FINDINGS	19
CONCLUSIONS	22
BIBLIOGRAPHY	23
APPENDICES	28

LIST OF FIGURES

FIGURE 1. Gallows Characters in Currier's Alphabet and Voynichese Equivalents	15
FIGURE 2. First Five Lines of Source and Sample Texts	18
FIGURE 3. Standard Deviation in Source and Sample Texts ..	19

APPENDICES

APPENDIX A. Logarithmic Distribution of All Samples	28
APPENDIX B. NO A Logarithmic Distribution	29
APPENDIX C. NO B Logarithmic Distribution	30
APPENDIX D. NO BV Logarithmic Distribution	31
APPENDIX E. NO PF Logarithmic Distribution	32
APPENDIX F. NO BVPF Logarithmic Distribution	33
APPENDIX G. W To S Logarithmic Distribution	34
APPENDIX H. WQXY To S Logarithmic Distribution	35
APPENDIX I. NO BV, WY To S Logarithmic Distribution	36
APPENDIX J. ALLGONE Logarithmic Distribution	37

INTRODUCTION

The Voynich manuscript is a vellum-bound book located in Yale University's Beinecke Rare Book and Manuscript Library. It is at least 400 years old and may have been written as early as the thirteenth century. The manuscript is of interest principally because the text is encrypted in a code that has sent scholars and code breakers away in defeat for nearly 100 years. The Voynich manuscript is not written in any known language - even the character set is unique to the corpus. It is, for many, an irresistible puzzle. The problem of the manuscript has aroused interest from many disparate groups - medievalists, linguists, computer scientists, and cryptographers. Members of the US intelligence community have been involved with breaking the Voynich cipher unofficially since the 1940's (Reeds, 1995). Little is known of the provenance and authorship of the manuscript, and theories about its content and origins abound (D'Imperio, 1977, Landini & Zandbergen, 1998).

Efforts have been made to examine the text using the assumption that it is a coherent document written in a natural language. These attempts have ruled out simple substitution ciphers and other elementary encryption schemes. Word length, character and term frequency, and comma counts all yield results that hint at something more than gibberish. Landini has demonstrated that the manuscript satisfies the rank-frequency,

number-frequency, length-frequency and length-rank laws for natural language text (Landini, 1997).

Since the translation effort is a massive undertaking involving many researchers, this study posed a small but concrete question whose answer is intended to chip away at the mystery of the manuscript.

Looking at the actual text of the Voynich manuscript presents a useful avenue of approach. Since the character set is entirely unique (with obviously unknown meaning), there is a great deal of speculation about the position, composition, and formatting of characters, words, and word strings. One possibility raised in discussion by Voynich scholars is that a specific and frequently occurring set of characters (the "gallows" variants) represent textual metadata - perhaps nulls or word breaks (Grant, 2000). Since this is a subject that can be effectively investigated quantitatively, it became the focus of this study.

Examining the manuscript with the assumption that the gallows characters are non-textual shed light upon the underlying structure. By eliminating these eight characters in total, and in various combinations, it was possible to compare altered texts to the intact original and examine them statistically. Observed differences and similarities demonstrated the semantic importance of the gallows characters to the manuscript as a whole.

LITERATURE REVIEW

This review examines scholarly and pseudo-scholarly commentary related directly to the Voynich manuscript itself. It would be possible to expand the review into fields as diverse as philology, botany, cryptography, and medievalism, but a clear decision was made to narrow the scope to canonical Voynich references. Many redundant articles (those refuting Newbold, for example, of which there are many) have been omitted.

It is noteworthy that research related to the Voynich manuscript has more or less ruined the careers of several prominent scholars (Grossman, 1999), and is widely regarded as a fringe avocation.

History of the Voynich manuscript

The facts surrounding the provenance of the Voynich manuscript are few. We know, for example, that it was sold to Rudolph II of Bohemia between 1584 and 1588, for the grand sum of 600 ducats - but we do not know who sold it to him. We know a few hands through which it passed before coming to rest in the Jesuit library of the Villa Mandragone, outside Rome, for 250 years - but there are large gaps in the ownership history. And we know that rare book collector Wilfrid Voynich snatched it up in 1912. (D'Imperio, 1977).

The manuscript itself is richly illustrated (the author seems to have had a fondness for sketching nude women) and written on good quality vellum. It measures six by nine inches. It originally consisted of at least 116 folios, but only 104 have survived. The manuscript is clearly divided into topical sections ("herbal", "astrological", etc), which can be ascertained by the nature of the illustrations (Landini & Zandbergen, 1998). The text is, of course, entirely unreadable.

Unsuccessful Attempts

The first scholar to seriously examine the Voynich manuscript was William Romaine Newbold, Professor of Philosophy the University of Pennsylvania. Newbold announced his successful decipherment of the manuscript in 1921. His translation was sensational - Newbold claimed that the manuscript was authored by the thirteenth century polymath Roger Bacon and contained miraculous accounts of Bacon's discoveries (Grossman, 1999). He determined that the cipher was composed of "microscopic shorthand signs" intermixed with a very subjective system of translation few but Newbold could repeat (Manly, 1931). His claims were at first accepted by the scholarly community, but later savagely ridiculed. His was the first career wrecked (albeit posthumously) by the Voynich manuscript. Bennett (1976, p.187) writes, "The works by Newbold . . . especially indicate the dangers of an ambiguous decoding method coupled with a vivid

imagination regarding the picture content." It is worth noting that Newbold, a Roger Bacon enthusiast to begin with, saw exactly what he wanted to see in the Voynich manuscript.

The other prominent figure brought low by his obsession with the Voynich manuscript was Yale Medieval Philosophy Professor Robert S. Brumbaugh, who announced his own breakthrough in interpretation (Brumbaugh, 1974). While criticizing Newbold's failed attempt, Brumbaugh fell into the same trap. His method considered the text "an artificial language, based on Latin, but not very firmly based there..." Brumbaugh stated that the text was (conveniently) "phonetically impressionistic" (1975, p.354). It is astonishing that he could completely miss the lack of rigor associated with the material he was publishing - it is clear to the detached scholar that Brumbaugh, like Newbold, let his enthusiasms get in the way of his scholarship.

Others have "solved" the riddle of the Voynich manuscript over time, attributing it variously to Ukrainian Khazars (Stojko 1978), English mystic Anthony Askham (Strong, 1945), or interpreting it as "a liturgical manual for the Endura rite of the Cathari heresy, the cult of Isis" (Levitov, 1987). Others have argued that the Voynich manuscript is simply gibberish (Williams, 1999) or a deliberate forgery by Wilfrid Voynich (Barlow, 1986).

Toward a Solution

William F. Friedman, beginning in 1944, first undertook a systematic analysis of the Voynich manuscript, using the computational tools of cryptanalysis. Friedman, one of the most famous American cryptographers of the Second World War, organized a study group to investigate the manuscript on an extracurricular basis (Reeds 1995). They developed the first transcription alphabet to make Voynich characters machine-readable. Using an early RCA computer and punch cards, the study group performed some rudimentary analyses with the assumption that the Voynich manuscript was a standard cipher text. The results of this investigation are not entirely known, but it is thought that many of the records - and the original IBM punch card set - are somewhere in the National Cryptologic Museum in Fort Meade, Maryland (Reeds, private communication, 10 Oct. 2000).

Others have looked at the Voynich manuscript through the lens of a scholarly specialty. Hugh O'Neill (1944), a botanist, identified several plants in the botanical section of the manuscript, notably the plant that dominates folio 93 recto - *Helianthus annuus*, the common sunflower. This identification, if correct, antedates the manuscript to 1493 at the earliest - the sunflower was not seen in Europe until brought back by Columbus.

In the late sixties and early seventies, Prescott Currier advanced the state of Voynich research with an important discovery. Currier proved statistically that there are two

distinct "hands" in the manuscript, each with a distinct subset of the Voynich script, representing multiple scribes (Currier 1975). He also demonstrated mathematically that, as in natural languages, lines are functional entities in the text. The importance of these observations cannot be emphasized enough. Most importantly, multiple authorship tends to rule out "grapholalia", or meaningless text. It is clear that more than one person copied the Voynich manuscript from an earlier source, and Currier identified the idiosyncrasies of the individual copyists. The concept of multiple "languages" has influenced critical thought on the nature of the Voynich manuscript.

In the 1970's, Mary D'Imperio, a mathematician and NSA consultant, was instrumental in encouraging the rigorous scientific examination of the manuscript. Her The Voynich Manuscript--An Elegant Enigma (1978) provides a thorough and accurate picture of the state of research up until that time, building on a series of articles she wrote for Manuscript (1977). I will discuss it at length because it is generally accepted as the starting point for all serious Voynich manuscript research.

D'Imperio begins with a brief history of the manuscript, including the known provenance and ownership history. This takes all of two pages, so scant is our knowledge of the manuscript's past. She then dives into a survey of "methods of attack" - possible avenues of approach in the decipherment effort, including content analysis of the drawings and cryptanalytic attacks on the text itself.

An Elegant Enigma covers the failed decipherment efforts of Newbold, Brumbaugh, and others gently, and then reviews the more serious efforts of her mid-seventies contemporaries, including Prescott Currier, with whom D'Imperio worked closely. Her book could be seen as a continuation and expansion of both Currier and Friedman's work on the Voynich manuscript. This "pedigree" represented the most authoritative decipherment effort up until the late nineties.

Three large sections cover collateral research in the areas of medieval iconography, secret languages, and early herbal manuscripts thought to be contemporary with the Voynich manuscript. D'Imperio concludes with suggestions for further research, some of which have been acted on (Stallings 1998, Zandbergen 1997, Guy 1991). The manuscripts current owner, Yale University, has resisted others, such as a paleographic examination of the text itself. Because of this, we do not know the age of the manuscript (the vellum could be easily carbon dated, giving us a "not before the calf died" approximate date of origination).

Much of the work related to the Voynich manuscript has taken place in the realm of cryptography. One of the biggest stumbling blocks to cryptanalytic examination of the manuscript in the late seventies was the lack of a consistent machine-readable draft of the text, something that was remedied in the eighties and nineties. This allowed a detailed analysis of the Voynich manuscript using information retrieval techniques and

large-scale data manipulation, which has led to some interesting conclusions.

With so opaque an artifact as the Voynich manuscript, scholarship has taken a subtractive approach - we are slowly learning what it is not, rather than gaining any insight into what it is.

O'Neill (1944), if correct, placed the manuscript after 1493. Currier (1976) established that it is statistically not gibberish. Landini (1997) and Landini and Zandbergen (1998) demonstrated that the Voynich manuscript exhibits lower entropy than comparable natural language texts, possibly indicating a similarity to sixteenth-century artificial languages. Stallings (1998) investigated the roots of second-order entropy in the text, dismissing the possibility of a low-entropy language, like Hawaiian, as the plaintext (Stallings does not suppose Hawaiian to be a realistic possibility - he uses it as a real-world test for a polysyllabic, low-entropy plaintext, most likely an artificial language). Stallings also demonstrated conditions in which a cipher can return results similar to those exhibited by the Voynich manuscript. Perakh (1999), using letter serial correlation, independently confirmed that the Voynich manuscript is not a random or quasi-random collection of characters.

Current Research

The more contemporary studies (those of Landini, Zandbergen, Perakh, and Stallings as mentioned above) represent a "new wave" of Voynich manuscript research. Its features include independent cryptanalytic studies without peer review and open availability through the medium of the World Wide Web. While not methodologically rigorous, findings generated and shared in this method are open to review, commentary, replication, and criticism. They are the work of both professionals (like Jim Reeds, a mathematician and cryptanalyst), and talented amateurs. Each study has a point of contact with the Voynich manuscript that is relevant and familiar to the researcher, but does not necessarily build on previous work. The fact that legitimate inquiry into the Voynich manuscript is frowned upon in academia greatly hampers decipherment efforts, since research is necessarily done in free time and published sporadically, generally on the Web.

While "serious" Voynich manuscript research has been sidelined since the humiliation of Newbold and Brumbaugh, a community of scholars continues to do important work in an informal and cooperative forum. The connectivity scholars enjoy thanks to the Internet has accelerated progress in several potentially fruitful areas. A unified transcription alphabet, developed by the European Voynich Manuscript Transcription project (under the direction of Gabriel Landini) now seeks to

supercede the assorted alphabets individual researchers developed in the past. The entire text is available in ASCII form. Negotiations are underway with the Beinecke library to create a digital copy of the Voynich manuscript in meticulous detail (approximately 20 MB per folio page). An active, dedicated mailing list offers a forum for those interested in Voynichiana to share ideas, sources, and techniques.

METHODOLOGY

This study intended to answer a simple question: How does the elimination of gallows variants from the transcription set change the results of statistical queries on the Voynich manuscript? The data collection and analysis associated with this research project were carried out in Manning Hall, using the resources provided to all students in the School of Information and Library Science at the University of North Carolina, Chapel Hill.

It was hypothesized that the gallows variants in the Voynich manuscript alphabet are null characters, and that removing them will not have a statistically relevant impact on correlational power curves, such as those generated by the application of Spearman's rank correlation coefficient.

This study was designed to create samples that, despite various characters being completely removed, continued to strongly correlate with **CURRIER**, the base text derived from the

voynich.now file. It is possible that such correspondence would indicate the presence of null characters, whose removal did not affect the statistics of the modified text with statistical significance. Conversely, if the modified samples exhibit variation consistent with their rank and frequency within the manuscript, this is strong evidence that the characters are not null.

Sample texts that deviated from the **CURRIER** model and more closely resembled **QU'RAN** (an excerpt from the Holy Qu'ran) or **GENESIS** (a transcription of the Book of Genesis written in vulgate Latin) might indicate that the omissions in the sample illustrated a semantic relationship with a known language. Such correspondence would be tenuous, since a variety of causes could account for it, but the possibility of such a correlation was not dismissed, and samples were compared with **QU'RAN** and **GENESIS**.

The actual analysis was a straightforward application of Spearman's rank correlation coefficient to nine separate data samples, along with the source text and two natural language control files.

Data Collection

All the raw data used in this study was obtained from public FTP archives accessed via the World Wide Web.

The quantitative analysis was based on the voynich.now file, which is freely available on the World Wide Web (Gillogly,

2001). `voynich.now` is a machine-readable version of Prescott Currier's Voynich transcription, using his alphabetic coding scheme.

`Voynich.now` was chosen over a subset of the European Voynich Manuscript Transcription project transcription file, which was created and is maintained by Gabriel Landini, and which is also available on the World Wide Web (Landini et al, 1998). The EVMT transcription files consist of an error-checked compilation of earlier transcription files (principally the Currier and FSG files) using a clear interpretation of the Voynich manuscript character set that includes gallows characters as amalgams of individual symbols rather than pairs or ligatures, as has previously been assumed. The EVMT files are the most accurate available, and represent the best hope of a standardized alphabet for disparate researchers to adopt at this time. Although not perfect (like any Voynich manuscript transcription alphabet, it carries with it a set of assumptions about the underlying text), the EVMT alphabet (called EVA) was designed to be over-broad rather than over-narrow.

The Currier transcription, although older and arguably less accurate, was chosen because the gallows characters are discrete, rather than composed of groupings of characters. Using `voynich.now` made the sample texts less ambiguous and simpler to code and process. As an example, the same gallows character represented by "W" in Currier's alphabet is "cph" in EVA. Since the analysis was conducted on the complete manuscript (the sum

total of Voynich character information in existence) the more granular EVA character set was deemed unnecessary.

Data Analysis

The principal tools for analysis were the SPSS and TACT statistical software packages. TACT is freely distributed on the World Wide Web (Bradley et al, 1993), and the correlational data supplied by SPSS could be generated by hand, or by other hardware or software.

As controls, two known-language samples of near-identical size were analyzed along side the voynich.now base text and modified samples. Arabic, a language that contains a significant volume of null characters, was chosen as the first control. A romanized 140K sample of the Holy Qu'ran was used, since this, like the vulgate Latin bible Landini et al use as a benchmark (Landini, 1987), is likely to represent an Arabic contemporary of the Voynich manuscript with some verisimilitude. The second control was a sample from a vulgate Latin translation of the book of Genesis.

The Source Texts

Two texts were used as controls in this study. The first is a **GENESIS** (in ASCII format), a 163K text file. The second is

QU'RAN, written in the ASMO 708 (ISO-8859-6) encoding scheme, a 124K text file.









The Voynich file used for this study is the readily-available voynich.now version of Currier's transcription (identified hereafter as **CURRIER**), written in ASCII and encoded in Currier's version of the Voynich alphabet. The version used for analysis was 120K in size.

All three were standardized by removing comments and extraneous material. In the case of the Qu'ran, ISO-8859-6 characters were translated to arbitrary vanilla ASCII characters on a one-to-one basis, working through the Latin alphabet in the order the characters appeared in the text. The resulting file uses the Latin characters A-Z and a-k.

The Modifications of CURRIER

The **CURRIER** text file was modified in nine different ways, to investigate the nature and relevance of various characters within the Voynich manuscript. Two groups of characters were of particular interest: the "gallows" characters represented by B, V, P, and F in Currier's alphabet, and the "gallows ligatures" represented by W, Q, Y, and X.

FIGURE 1: Gallows Characters in Currier's Alphabet and Voynichese Equivalents

B	V	P	F	W	Q	Y	X
							

W, Q, Y, and X all appear in combination with another Voynichese symbol, represented in Currier's alphabet as S.

B, V, W, and Y are quite similar, each having only one "leg", while P, F, Q, and X all have two.

Given these facts, several possibilities present themselves. The **NO B** modification is intended to explore the possibility that each character in the gallows group is discrete by removing only the B, and not its one-legged analogs (V, W, and Y). Similarly, The **W to S** modification is intended to accomplish the same thing with an "overlaid" gallows character, converting only W to the underlying S.

The **NO A** modification serves as a checksum, since the A character is statistically similar to the gallows variations in frequency, but shares none of their characteristics in the Voynich manuscript.

The **NO BV** modification removes a one-legged pair of gallows but leaves W and Y, their "over S" analogs, in place. This version was intended for comparison with **NO PF**, which removed the two-legged gallows pair.

NO BV, WY to S explores the possibility that the gallows characters are linked. All one-legged versions are removed from this modification.

BVPF removes B, V, P, and F characters, while retaining the "overlaid" gallows analogs. There were 752 B, 3319 P, 202 V, and

5469 F instances, reducing the MS by 9722 characters, from 110,977 to 101,255.

The **WQYX to S** variant is the opposite of **BVPF**. It removes the gallows overlays and converts them to the underlying character, S. Assuming they are null, the semantic value of the underlying S would be intact. The result of this was the replacement of 146 W, 709 Q, 51 Y, and 561 X characters. The new total is 8064 S characters, up from 6,597. The MS length was obviously unchanged.

ALLGONE removes every gallows character from the text. W, Q, X, and Y are replaced with S, the underlying character, and B, V, P, and F are simply expunged.

Once all twelve text files were prepared, each was processed using TACT (Text Analysis and Computing Tools), a text-retrieval and analysis software package developed at the University of Toronto. TACT, which was designed for use with small groups of literary texts using western alphabets (Bradley et al, 1993), parsed the text and returned detailed information on frequency (rank, percentage, and number of words) as well as type and token information. TACT also generated thesauri, and word and character lists useful for further statistical analysis.

Assumptions and Limitations

This study assumes that a null is truly meaningless and not a blank space - that it is without meaning in the context of the

document formatting. Thus, when it is removed, the adjacent characters are truncated into a new, shorter word, rather than becoming two separate words.

The Arabic sample used in this study is modern and unvocalized, rather than classical, Arabic. The Currier transcription is incomplete and imperfect, and makes assumptions about the alphabet that may be entirely incorrect.

FIGURE 2: The First Five Lines of Source and Sample Texts

CURRIER (IN VOYNICHESE)

١٤٢٩٢ ٩٩٢٢ ١٢٢٢ ١٢٢٢ ١٢٢٢ ١٢٢٢ ١٢٢٢ ١٢٢٢ ١٢٢٢ ١٢٢٢
 ٢٢٢٢ ٢٢٢٢ ٢٢٢٢ ٢٢٢٢ ٢٢٢٢ ٢٢٢٢ ٢٢٢٢ ٢٢٢٢ ٢٢٢٢ ٢٢٢٢
 ٢٢٢٢ ٢٢٢٢ ٢٢٢٢ ٢٢٢٢ ٢٢٢٢ ٢٢٢٢ ٢٢٢٢ ٢٢٢٢ ٢٢٢٢ ٢٢٢٢
 [٢٢٢٢]٢٢٢٢ ٢٢٢٢٢٢ ٢٢٢٢٢٢ ٢٢٢٢٢٢ ٢٢٢٢٢٢ ٢٢٢٢٢٢ ٢٢٢٢٢٢ ٢٢٢٢٢٢
 ٢٢٢٢ ٢٢٢٢ ٢٢٢٢ ٢٢٢٢ ٢٢٢٢ ٢٢٢٢ ٢٢٢٢ ٢٢٢٢ ٢٢٢٢ ٢٢٢٢

CURRIER

VAS92.9FAE.AR.APAM.ZOE.ZOR9.QOR92.9.FOR.ZOE89.
 2OR9.XAR.O.R.9.FAN.ZPAM.ZAR.AR*.QAR.QAR.8AD.
 29AU.ZCF9.OR.9FAM.ZO8.QOAR9.Q*R.8ARAM.29.
 [082*]OM.OPCC9.OPCOR.2OEO9.Q*AR.8AM.OFAM.OE.OFAD.
 2AT.9.SCAR.QAM.WAR.YAM.

WQXY To S

VAS92.9FAE.AR.APAM.ZOE.ZOR9.SOR92.9.FOR.ZOE89.
 2OR9.SAR.O.R.9.FAN.ZPAM.ZAR.AR*.SAR.SAR.8AD.
 29AU.ZCF9.OR.9FAM.ZO8.SOAR9.S*R.8ARAM.29.
 [082*]OM.OPCC9.OPCOR.2OEO9.S*AR.8AM.OFAM.OE.OFAD.
 2AT.9.SCAR.SAM.SAR.SAM.

NO BVPF

AS92.9AE.AR.AAM.ZOE.ZOR9.QOR92.9.OR.ZOE89.
 2OR9.XAR.O.R.9.AN.ZAM.ZAR.AR*.QAR.QAR.8AD.
 29AU.ZC9.OR.9AM.ZO8.QOAR9.Q*R.8ARAM.29.
 [082*]OM.OCC9.OCOR.2OEO9.Q*AR.8AM.OAM.OE.OAD.
 2AT.9.SCAR.QAM.WAR.YAM.

ALLGONE

AS92.9AE.AR.AAM.ZOE.ZOR9.SOR92.9.OR.ZOE89.
 2OR9.SAR.O.R.9.AN.ZAM.ZAR.AR*.SAR.SAR.8AD.
 29AU.ZC9.OR.9AM.ZO8.SOAR9.S*R.8ARAM.29.
 [082*]OM.OCC9.OCOR.2OEO9.S*AR.8AM.OAM.OE.OAD.
 2AT.9.SCAR.SAM.SAR.SAM.

FINDINGS

Standard Deviation

Looking at Standard Deviation, with Latin and Arabic known-language texts included for comparison, produced results consistent with the hypothesis that gallows characters have meaning and are not null.

FIGURE X: Standard Deviation in Source and Sample Texts

TEXT	STANDARD DEVIATION
CARRIER	16.9245
W to S	17.0761
NO B	17.5329
NO A	17.7746
NO BV	17.7827
QU'RAN	17.8885
NO BV, WY to S	18.0343
WQYX to S	18.2988
NO PF	22.1239
NO BPVF	23.733
ALLGONE	26.0574
GENESIS	29.5734

Standard Deviation increases in every manipulation of the source text. The most marked increase occurs in those versions with character omissions (**NO BPVF** for example). The total conversion modification (**WQYX to S**) also shows a significant increase in Standard Deviation. It is likely that the high Standard Deviation resulting from total removal of all gallows characters (**ALLGONE**), when compared to the source text, reflects lexical importance in at least some of the gallows characters.

It is interesting to note that **NO BV** and **NO PF** returned quite different Standard Deviations.

Rank Correlation

The rank order of terms within the text samples may be significant. By looking at rank, rather than frequency, it is possible to differentiate between texts, looking for close pattern matches, and apply appropriate statistical measures.

The Spearman rank correlation coefficient (Spearman's rho) was used to examine differences in rank order between samples, as well as levels of significance.

The Spearman rank correlation coefficient is an outgrowth and expansion of the Pearson correlation coefficient, and is designed for use with ordinal data (Roscoe, 1969). Thus, it is an excellent tool for examining differences in rank.

To do this, the Voynich and related data was first organized from each text sample by frequency, re-ordering rank in the process. The Spearman rank correlation coefficient was then applied.

As might be expected, all the samples, when compared statistically with the source **CURRIER** text, returned scores well above the .01 level of significance. The lowest nonparametric correlation, **NO BV**, was 0.426, and they ranged as high as 0.69 (**NO BVPF**).

Performing a logarithmic transformation of the frequency data of the texts yielded a set of relevant graphs illustrating similarity and difference between samples (See Appendices A-J). Each sample text demonstrated consistent negative correlation.

The sample texts that most closely mimic the original **COURIER** are **NO B** and **W to S**, with **NO B** diverging only slightly (less than 1%) from the source text. **CURRIER** has 217 more unique words than **NO B**, but the samples are otherwise quite similar. The character B represents 1.44% of the text, and is the 13th most common letter in **CURRIER**. In contrast, the character A represents 3.98% of the text, and is 7th most common in **CURRIER**. **NO A**'s correspondence with **CURRIER** is not as close as **NO B**, which is most likely due to the relative importance of the letter in the Voynichese alphabet. This, in turn, argues against ascribing importance to the close correspondence between **NO B** and the source text.

The sample texts with the widest divergence from **COURIER** (excluding the known-language texts **GENESIS** and **QU'RAN**) are **NO PF** and **ALLGONE**. In the case of **ALLGONE**, 10.49% of the characters in the manuscript have been stripped away (every gallows variant). **NO PF** removes 5.86%, the two most frequent gallows variants. The wide variance between these samples and the source text suggests that the wholesale removal of the gallows characters has a profound impact on the underlying structure. This, in turn, points toward lexical significance for those characters.

CONCLUSIONS AND OPPORTUNITIES FOR FURTHER RESEARCH

This study provided evidence to support the conclusion that the gallows characters, individually and as a group, are not null characters. The elimination of gallows variants from the transcription set changed the results of statistical queries on the Voynich manuscript in ways that are consistent with value-laden characters of the same rank and frequency.

It was hypothesized that the gallows variants in the Voynich manuscript alphabet are null characters, and that removing them would not have a statistically relevant impact on correlational power curves. This did not prove true.

The Voynich manuscript bristles with untouched problems for the enthusiastic researcher. Several possibilities related to this study present themselves.

Most obviously, although the gallows variants present likely candidates for nulls, the rest of the Voynichese alphabet awaits a thorough analysis using the same methodology.

In addition, re-coding the sample texts with the assumption that gallows characters had significance to the document format could be instructive. If they represent word breaks, replicating this study with revised samples would illustrate this.

BIBLIOGRAPHY

Barlow, Michael. (October 1986). The Voynich Manuscript -- By Voynich? Cryptologia 10, 210-216.

Bennett, William Ralph. (1976). Scientific and Engineering Problem Solving with the Computer. Englewood Cliffs: Prentice-Hall.

Bradley, John et al. (1993). TACT [Computer Program]. Toronto, ON: University of Toronto TACT Group.

Brumbaugh, R. S. (1974). Botany and the Voynich 'Roger Bacon' Manuscript Once More. Speculum 49, 546-48.

Brumbaugh, R. S. (1975). The Solution of the Voynich 'Roger Bacon' Cipher. Yale Library Gazette 49, 347-55.

Currier, Prescott A. (1976). Some Important New Statistical Findings.

Retrieved 20 October, 2000, from the World Wide Web:

<ftp://ftp.rand.org/pub/voynich/currier.paper>

D'Imperio, M. E. (Spring 1977). The Voynich Manuscript: A Scholarly Mystery. Manuscripts 29,(2), 85-93.

D'Imperio, M. E. (Summer 1977). The Voynich Manuscript: A Scholarly Mystery. Manuscripts 29,(3), 161-173.

D'Imperio, M. E. (Winter 1978). The Voynich Manuscript: A Scholarly Mystery. Manuscripts 30,(1), 34-48.

D'Imperio, M. E. (1978). The Voynich Manuscript--An Elegant Enigma. Laguna Hills: Aegean Park Press.

Gillogly, Jim, et al. (2001) voynich.now [ASCII computer file]. Retrieved April 16, 2001 on the World Wide Web:
<ftp://ftp.rand.org/pub/voynich>

Grant, Bruce (bgrant@mail.msen.com). (10 June 2000). "Re: Curious coincidence." E-mail to Voynich manuscript mailing list (voynich@rand.org).

Grossman, Lev. (April 1999). When words fail: The struggle to decipher the world's most difficult book. Lingua Franca 9,(3), 9-15.

Guy, J. B. M. (1991). Statistical Properties of Two Folios of the Voynich Manuscript. Cryptologia 15, 207-218.

Landini, Gabriel. (1996) European Voynich Manuscript Transcription. Retrieved April 16, 2001 on the World Wide Web:
<http://web.bham.ac.uk/G.Landini/evmt/rules.htm>

Landini, Gabriel., and Rene Zandbergen. (July 1998). A Well-kept Secret of Medieval Science: the Voynich Manuscript. Aesculapius 18, 77-82.

Landini, Gabriel. (1997, revised 2000). Zipf's laws in the Voynich Manuscript. Retrieved 20 October, 2000, from the World Wide Web:
<http://web.bham.ac.uk/G.Landini/evmt/zipf.htm>

leb.net. (2001) alquran.txt [ASCII computer file]. Retrieved April 16, 2001 on the World Wide Web:
<http://leb.net/qalam/islam/quran/alquran.txt>

Levitov, Leo. (1987) Solution of the Voynich Manuscript: A Liturgical Manual for the Endura Rite of the Cathari Heresy, the Cult of Isis. Laguna Hills, California: Aegean Park Press.

Manly, John M. (July 1931). Roger Bacon and the Voynich Manuscript. Speculum 6, 345-91.

McKenna, Terence K. (1991). The Archaic Revival. San Francisco: HarperSanFrancisco.

O'Neill, Hugh. (1944). Botanical Observations on the Voynich MS. Speculum 19, 126.

Perakh, Mark. (1999). Application Of The Letter Serial Correlation Test To The Voynich Manuscript. Retrieved 20 October, 2000, from the World Wide Web:

<http://www.nctimes.net/~mark/Texts/voynich1.htm>

Reeds, Jim. (1995). William F. Friedman's Transcription of the Voynich Manuscript. Cryptologia 19, 1-23.

Roscoe, John T. (1969). Fundamental Research Statistics for the Behavioral Sciences. New York: Holt, Rinehart and Winston.

Stallings, Dennis. (1998). Understanding the Second-Order Entropies of Voynich Text. Retrieved 20 October, 2000, from the World Wide Web:

<http://www2.micro-net.com/~ixohoxi/voy/mbpaper.htm>

std.com. (2001) genesis.txt [ASCII computer file]. Retrieved April 16, 2001 on the World Wide Web:

<ftp://ftp.std.com/WWW/obi/Religion/Vulgate>

Strong, L. C. (15 June 1945). Anthony Askham, the author of the Voynich Manuscript. Science 101, 608-9.

Stojko, John. (1978). Excerpts from Letters to God's Eye: The Voynich Manuscript for the first time deciphered and translated into English. Retrieved 20 October, 2000, from the World Wide Web: <http://home.att.net/~oko/voynich.htm>

Voynich mailing list. Available via electronic mail:
voynich@rand.org.

Williams, Robert L. (October 1999). A Note on the Voynich Manuscript. Cryptologia 23, 305-309.

Zandbergen, Rene. (1997). Currier A and B: two different languages? Retrieved 20 October, 2000, from the World Wide Web:
<http://www.bham.ac.uk/G.Landini/evmt/lang.htm>