No News about the Voynich Manuscript?

(Paper based on an invited talk at the 2019 HistoCrypt meeting, 23-25 June 2019, Mons, Belgium)

René Zandbergen,

02/07/2019

Abstract

The Voynich Manuscript (MS) is an unreadable manuscript from the 15th century that has defied all attempts of translation, and has become one of the main challenges for historical cryptography. The MS was acquired by the book dealer Wilfrid Voynich in 1912. No acceptable translation of the MS has been produced in the more than 100 years since then. At the same time, however, in the last decade or so there have been regular reports in the media that the mystery of the Voynich MS has been solved by someone. So what is really the state of progress with respect to the Voynich MS?

This paper first looks at some of the things that we have learned about the MS in the last few years (the "News"). Then it briefly lists a few of the acclaimed but unsuccessful translations or 'solutions' of the text ("No News") and finally it looks at a few of the main reasons why all of these recent attempts have been unsuccessful ("Why No News"?).

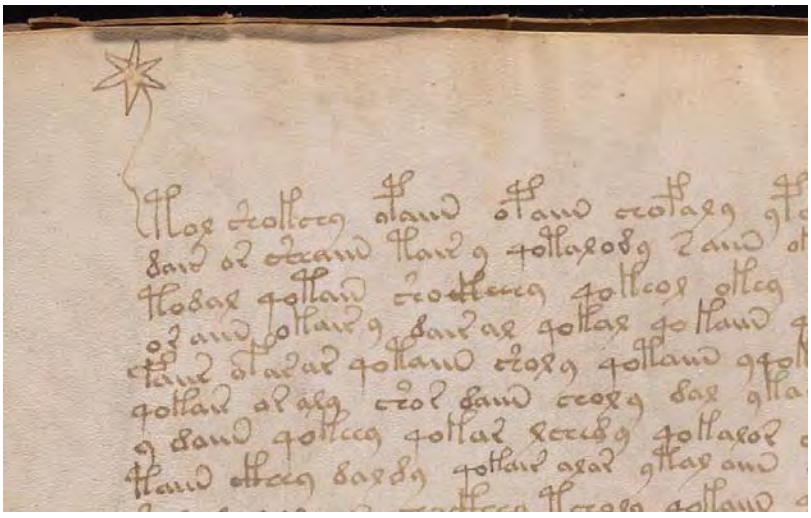


Figure 1: example of the Voynich MS writing

Brief Introduction

The Voynich MS is a parchment manuscript written in the 15th century. It has well over 200 pages, mostly with drawings of plants, constellations and strange systems of tubes populated by small female figures. Essentially all pages are filled with text in a script that is not known from any other document.

This text appears to consist of words and uses some 25-30 different symbols. An example is show above (Figure 1). For many more details about the MS, the reader is referred to the website of the author: http://www.voynich.nu/ .

The MS is now preserved at the Beinecke Rare Book and Manuscript Library of Yale University, New Haven (CT), henceforth simply the Beinecke Library.

PART 1: Some news related to the Voynich MS

Forensic investigations of the MS in 2009

The most important 'recent' news about the Voynich MS dates from 2009 and is generally known by now. In that year, the Beinecke Library agreed to submit the Voynich MS to forensic investigations, including radio-carbon dating of the parchment and chemical investigation of the inks and paints used in the MS.

The radio-carbon dating showed that the parchment dates from 1405-1438 with 95% probability, while the analysis of the inks and paints showed that all components used were commonly used in medieval manuscripts for several centuries. Specifically, no unusual or 'suspect' component could be identified, unlike for example the case of the Vinland map. The results have been presented at several workshops and symposia [1].

Voynich MS Exhibition and Workshop in 2014

A workshop related to the Voynich MS was co-organised by the Folger Shakespeare Library in Washington DC and the Beinecke Library, in November 2014. On this occasion, the MS was closely inspected for a few hours by several MS conservators from important US libraries, by an art historian and by a historian of medicine with expertise in herbal manuscripts [2].

Apart from these inspections, the MS was further submitted to additional forensic tests. The main findings from this event are:

- The parchment of the MS is made of calf skin, as determined by the University of York (UK) using protein sequencing.
- This parchment is not of very high quality, but it was very well prepared, to the extent that it is hardly possible to distinguish which is the flesh side and which the skin side of the parchment.
- The present cover is made of goat skin.

---

[1] For example: Zyats, Hodgins and Barabe, The Mysterious Voynich Manuscript: Collaboration Yields New Insights. American Institutes for Conservation's 40th Annual Meeting, May 8–11, 2012, Albuquerque, New Mexico. Details of the radio-carbon dating procedure may be found here: http://www.voynich.nu/extra/carbon.html .

[2] Their names may be found in Clemens, Raymond (ed.), with introduction by Deborah Harkness: The Voynich Manuscript, Yale University Press, New Haven and London, 2016 (Acknowledgments section).

- The present cover is not the original cover. An earlier cover of the MS had wooden boards covered by tanned leather.
- The stitching/sewing of the binding is very old, perhaps even 15th century.
- The MS lacks yellow paint, which is very unusual for a herbal MS.
- This apparent lack of yellow paint is caused by the use of organic colorants that have faded over the ages.
- The clothes worn by the figures in the zodiac are typical for the 1420's and would be unknown even a few decades later.
- Additional forensic tests confirmed the 2009 results. No unusual chemical constituents were found in the inks and paints.
- Multi-spectral imaging revealed no instance of erased text, beyond the entries already known on the first folio of the MS.

During the workshop, two publications about the Voynich MS were announced, which have both appeared in the meantime. The first is a very affordable photo-facsimile of the MS with several essays [3]. The second is a high-quality facsimile edition, produced by the Spanish company Siloé in Burgos [4].

Further progress related to the history of the MS

The history of the MS is summarised visually in Figure 2 below.

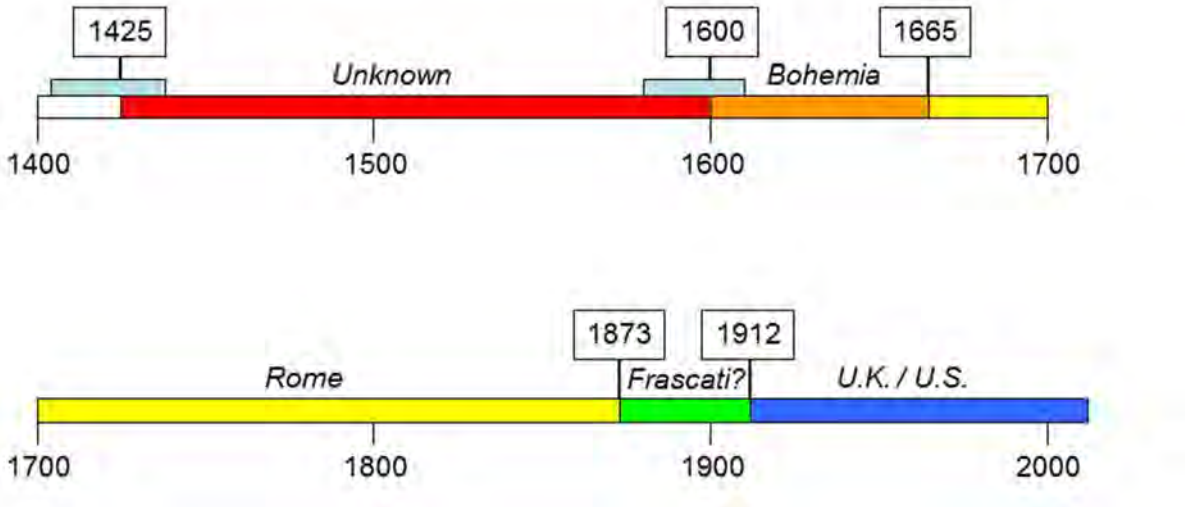


Figure 2: timeline of the (partially) known whereabouts of the Voynich MS

Nothing is known about the whereabouts of the MS before around 1600, when it was acquired by Emperor Rudolf II who resided in Prague. Then, until 1665 it was in Bohemia and we have several references to the MS until the year 1665 when it was sent to the Jesuit Athanasius Kircher in Rome.

[3] Clemens (ed.), 2016, see footnote 2.
[4] The first copies were presented on 10 November 2017 in Burgos.

Then follows another period of which we have no news. This period is even longer than the initial unknown period. We do know that the MS received its present cover during this time, in the Collegium Romanum library of the Society of Jesus in Rome. Then, latest in 1873, the MS was part of a collection that was hidden by the Jesuits in order to avoid confiscation by the state. Tradition says that the MS was kept in the Villa Mondragone in Frascati, until it was sold to Voynich in 1912. After that, he kept it in London and later in the US, where it is still today. There is some recent progress in three areas, which may be summarised in reverse chronology.

The highly mysterious conditions under which Voynich acquired the MS are gradually being unravelled:

- The 'discovery' of the MS by Voynich, in chests in a castle somewhere, is only a cover story. The MS was never lost.
- He was offered an opportunity to buy some old Jesuit manuscripts that had been hidden. He had to maintain absolute secrecy (and therefore needed the above cover story).
- The sale was from the Society of Jesus to the Vatican. The involvement of a private dealer seems to be linked to an attempt to return at least one valuable manuscript from the library of Mathias Corvinus to Hungary. (This did not work out – it was sold to Pierpont Morgan in the US).
- It is not at all certain that these manuscripts were ever in Villa Mondragone.

The oldest reference to the Voynich MS

The oldest positively identified reference to the Voynich MS was found already many years ago by the Czech historian Josef Smolka. It was digitised in 2016 and transcribed anew and translated by Philip Neal [5]. It is a letter from Athanasius Kircher to the Jesuit mathematician Theodor Moretus in Prague and it is dated 12 March 1639, see Figure 3. At this time the MS was owned by Georgius Barschius.

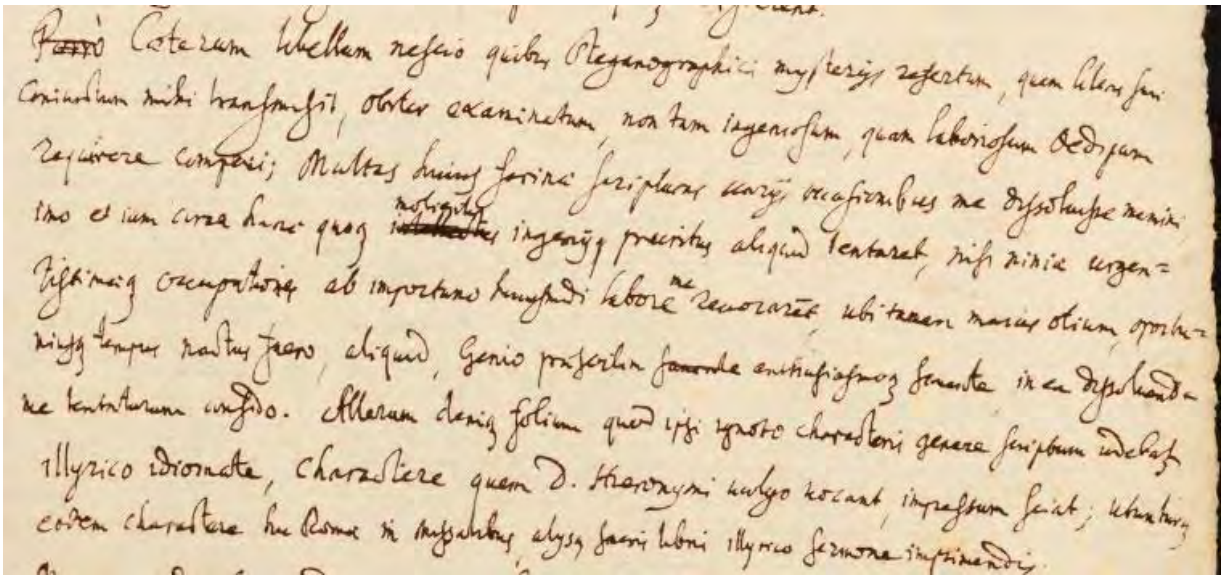


Figure 3: part of the 1639 letter from Kircher to Moretus

[5] For a full transcription and translation, see: http://www.voynich.nu/letters.html .

It says how Kircher had not yet been able to translate the mysterious writing that he had received from Moretus. It did not seem to be too difficult but would more be a matter of time. He might get back to it later. This must have prompted the well-known letter from Barschius to Kircher of 21 April 1639.

Who sold the Voynich MS to Rudolf II?

The sale to Rudolf II is not absolutely certain, but it was recorded by very reliable sources. It would be of great interest to identify the seller, as this would provide a new trail further back in time. In literature, one very often reads that the seller would have been John Dee, but this identification was based on a hypothesis of Wilfrid Voynich himself, and can be rejected [6].

In 2012 I suggested an alternative name, which was based on searches in Rudolf's account books. More recently I found further hints that would seem to strengthen this possibility. This topic is now being followed up by a trained historian who has made some very interesting discoveries. This work is not yet completed, but it is likely to be published by him later this year. Therefore, I will have to refrain from providing more details about this interesting work in progress.

PART 2: Some false news related to the Voynich MS

The mystery of the Voynich MS has finally been solved!

No, it hasn't, but similar headlines are appearing more and more regularly mostly in on-line media. A few examples of this may be cited.

- In 2004 it was 'proved' to be a meaningless fake by Gordon Rugg.
- In 2014 it was 'solved' by Stephen Bax.
    - In reality, he presented tentative translations for a dozen or so words.
- In 2017 the Times Literary Supplement wrote that Nicholas Gibbs had solved it.
    - This was subsequently rejected in later news publications.
- In 2018 it was announced that Artificial Intelligence had solved the Voynich MS. This referred to a paper by Hauer and Kondrak, suggesting that the text was anagrammed Hebrew.
    - This was subsequently rejected in later news publications.
- In May 2019 it was reported that G. Cheshire of Bristol University had solved the Voynich MS.
    - This theory was subsequently demolished in later news items.

In reality, many more proposed solutions do not even make the press, and they are all wrong, but why?

Having spoken with many of the people involved, I can point to a few common issues. With only a little bit of exaggeration:

- They all believe they are doing something completely new …

[6] This point is addressed in some detail in: Zandbergen, René and Rafał Prinke: The Voynich MS in Rudolfine Prague. In: Purš and Karpenko (2016), pp. 297-314.

- but in reality they are all doing the same thing …
- based on a number of assumptions (often without being aware),
- asking the wrong questions, assuming it is a cipher.
- The general working assumption seems to be that it is just a matter of finding the character conversion table and identifying the plain text language.

(Examples of people who are doing very different things are Rugg (2004) [7] and Timm and Schinner (2019) [8]).

There are many questions related to the Voynich MS text to which we don't yet have the answer, for example:

1. Is the text in the MS meaningful at all?
2. If so, was it generated using a reversible method?
3. Are the letter groups in the MS equivalent to plain text words?
4. If all three points are true, was the method similar to simple substitution? (Note that "writing in an invented alphabet" was one of the common historical encryption methods).

Many people tend to assume four times a "yes". However, there are other possibilities, again for example:

- This is someone's invented language …
    - which is meaningful only to him;
    - or not meaningful at all.
- It is a magical text. (Such texts also exist in readable manuscripts and appear meaningless).
- It just looks like language:
    - It was intended to make the owner look like a very wise person.

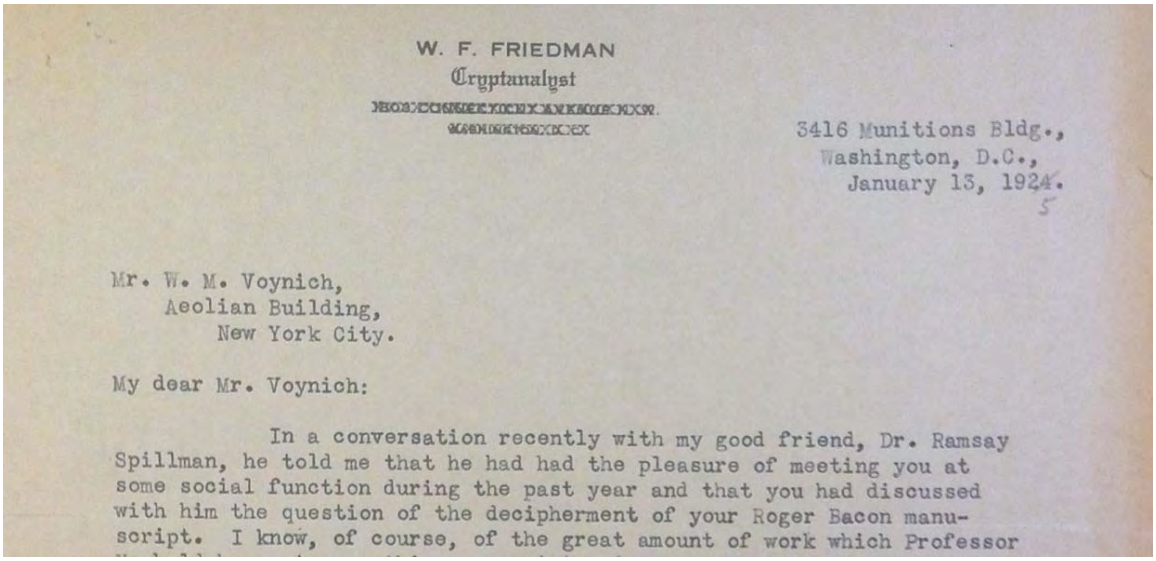
So, is the text of the MS really a cipher text? William Friedman, who is often considered the world's greatest cryptanalyst, spent a very long time studying the MS, from his first letter to Voynich in 1925 (see Figure 4) till his death in 1969. After all these years, he presented his opinion about the Voynich MS text in the form of an anagram, which he then resolved as follows:

The Voynich MSS was an early attempt to construct an artificial or universal language of the a priori type – Friedman.

That means that after 44 years of study, in which he set up at least two working groups to collectively analyse the text, he did not consider it to be a cipher.

[7] Gordon Rugg (2004) AN ELEGANT HOAX? A POSSIBLE SOLUTION TO THE VOYNICH MANUSCRIPT, Cryptologia, 28:1, 31-46

[8] Torsten Timm & Andreas Schinner (2019) A possible generating algorithm of the Voynich manuscript, Cryptologia



W. F. FRIEDMAN
Cryptanalyst

3416 Munitions Bldg.,
Washington, D.C.,
January 13, 1924.

Mr. W. M. Voynich,
Aeolian Building,
New York City.

My dear Mr. Voynich:

In a conversation recently with my good friend, Dr. Ramsay Spillman, he told me that he had had the pleasure of meeting you at some social function during the past year and that you had discussed with him the question of the decipherment of your Roger Bacon manuscript. I know, of course, of the great amount of work which Professor

Figure 4: the start of Friedman's (probably) first letter related to the Voynich MS.

This question can also be addressed in a very different way, by visually comparing the Voynich MS text to known early 15th century cipher texts, and to texts in unknown writing (as it would appear to someone not familiar with these examples). This is shown in Figure 5 further below. While this is of course very qualitative, the difference in appearance between known cipher texts and the Voynich MS text is quite clear. The main point of this is that it is dangerous to just assume that the Voynich MS text is a cipher text.

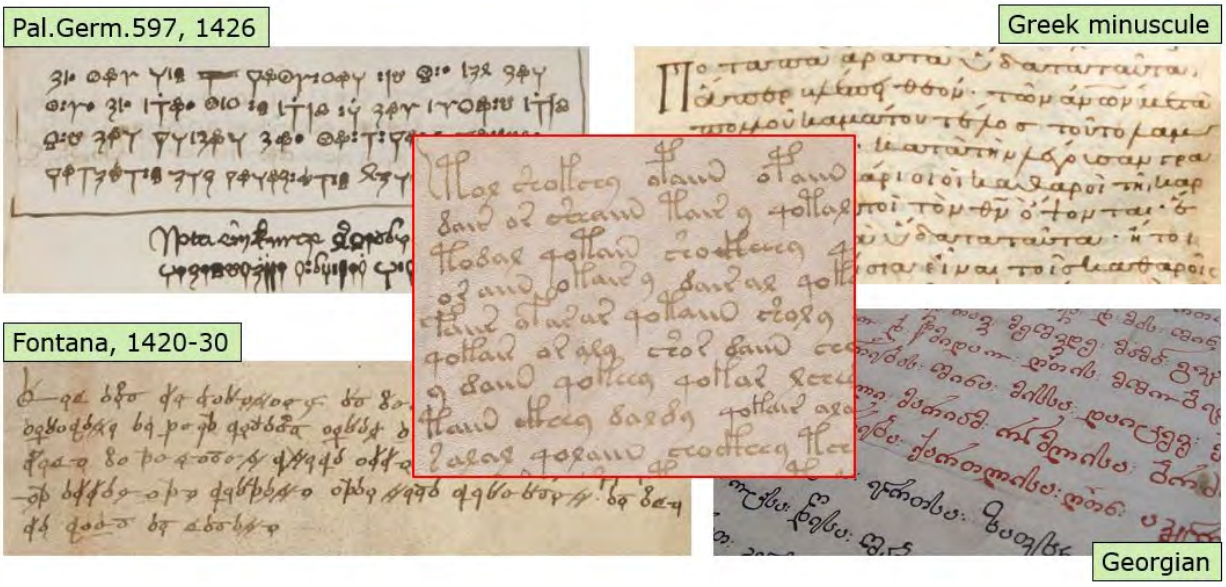


Figure 5: comparing the Voynich MS text to known cipher texts (left) and to unknown writing (right)

PART 3: Why everyone is getting it wrong

Asking the right question(s)

The standard question about the Voynich MS is: “What does it say?” This question may not have an answer. It may not say anything. It also leads people to working “backwards”, as follows:

- Take an electronic representation of the text;
- Manipulate it in order to arrive at some sort of “meaningful” plain text.

The right question should be: “How was it done?”, because this question definitely has an answer. It was most certainly ‘done’ one way or another, also if the text is meaningless. Looking for the answer to this question should lead to an attempt to reproduce this process, and hopefully to finding out the reverse process (if there is any).

The danger of working backwards may be demonstrated in the following. For this purpose it is first necessary to briefly introduce the character set used in the Voynich MS text:

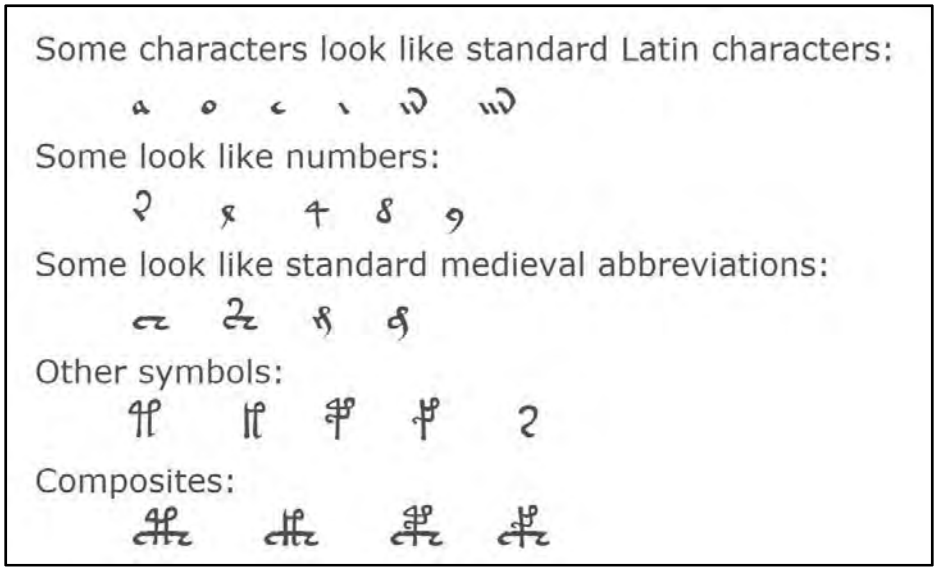


Figure 6: the set of most common characters used in the Voynich MS

Table 1 further below represents one proposed solution to the Voynich MS, by providing a translation from Voynich characters to plain text characters or groups of characters. To go from Voynich text to plain text is straightforward, but one has to imagine that the creator of the MS must have used the inverse of this table. It will be clear that in many instances this is ambiguous. Furthermore, many plain text characters are missing, namely b, c, g, h, j, k, p, v, w, x, y and z. Since the plain text language proposed by the originator of this ‘solution’ is Latin, some of these omissions cannot be explained. Furthermore, it will be clear from the table that a letter ‘s’ in the plain text could only be converted into Voynich text if it was preceded by an ‘e’ or an ‘o’.

In summary: the ‘encoding’ process from Latin to Voynich text is impossible.

Table 1: one proposed character assignment table from Voynich characters to plain text

Voynich	Latin		Voynich	Latin		Voynich	Latin
	-eus			i			te , ta
	-eum			um , ul, ur			eo
	m , n			o			f
	em			re , er			ad
	a			ex			d
	t			os			oe
	qu			e			ea
	r , l			te			ai , ea

The main unusual features of the Voynich MS text

The main reason for the many failed attempts is that people are not sufficiently aware of a number of unusual features or properties of the Voynich MS text. There are many, but the following three may be highlighted:

1. Some of the characters only appear in very specific environments:
   - at ends of words / ends of lines only;
   - mainly at the start of a paragraph;
   - almost exclusively top lines of paragraphs.
2. The vast majority of words exhibit a 'word structure':
   - This was first proposed by John Tiltman.
   - More elaborate patterns have been proposed by different researchers.
3. Entropy values for the text are anomalously low:
   - This is clearly related to the above-mentioned word structure, and it will be the main topic for the remainder of this paper.

To demonstrate what happens if one ignores the first point, we may look at the character assignment table (again 'backwards', i.e. from Voynich text to plain text) of the recent proposed solution by Gerard Cheshire [9]. It is given in Table 2 below. The problematic assignments are highlighted in yellow.

[9] Cheshire, Gerard: The Language and Writing System of MS408 (Voynich) Explained. Romance Studies (2019)

Table 2: character assignments by Gerard Cheshire in his 2019 'solution'

Voynich	Latin	Voynich	Latin	Voynich	Latin
	a		l		r
	a		m		s
	ais		ele		s
	aus		eme		ai , ea
	ae		n		t
	d		o		sa
	e		p		ta
	ee		qu		u
	(long) e		epe		v
	i		eque		

In the Voynich MS, the character assigned to Latin ‘d’ always appears at the start of a word, and it is essentially always followed by the character that looks like ‘o’. This means that in the plain text, the letter ‘d’ can only appear at the start of a word and is always followed by ‘o’. This is not a realistic behaviour of any Romance language (the proposed plain text language by Cheshire is “proto-Romance”).

The other four highlighted characters have the property that in the great majority of cases they appear in the top lines of paragraphs. This would mean that in the plain text words including a ‘p’ or a ‘q’ would almost only appear in the top lines of paragraphs, which is completely unrealistic.

Word patterns

The most conspicuous feature of the Voynich MS text is that there is some order or pattern defining which characters can appear in which positions in the words. It is not possible to treat this complicated topic adequately in this short paper, so two examples will have to suffice. The first example was found by John Tiltman, the famous British code breaker and collaborator of William Friedman. He noticed that the majority of words can be split into two parts which he called roots and suffixes. Each of the roots in Table 3 below can be combined with each of the suffixes, resulting in a valid Voynich word.

Table 3: Tiltman's split of words into 'roots' and 'suffixes'

Roots	Suffixes

A different approach was presented by Voynich researcher Michael Roe in the 1990's. It is shown in Figure 7 below. Each path through the structure, from left to right, results in a valid Voynich word.

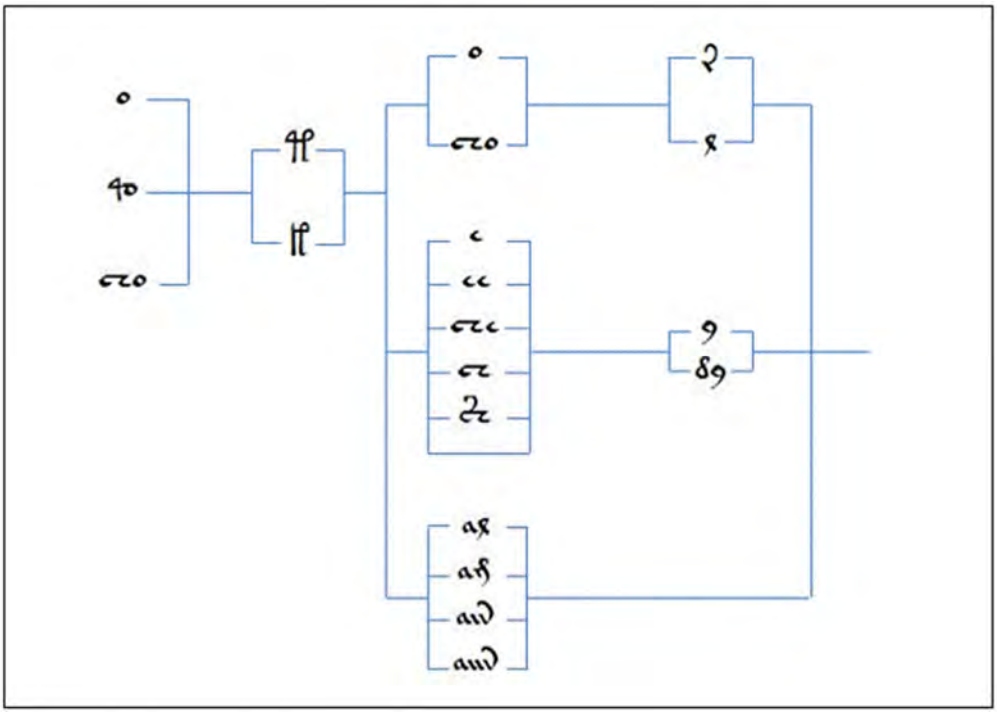


Figure 7: Mike Roe's pattern of common words

From both examples it will be clear that the text behaves very differently from 'normal' languages like Latin, Italian, German or English.

The anomalous entropy of the Voynich MS text

Entropy is a single number that characterises a probability distribution (also known as a partition). This value is maximum if all probabilities are equal. It is smaller in case the distribution is skewed. Entropy values can be computed and compared for different properties of a text:

- Character frequencies;
- Character combination frequencies;
- Word frequencies.

The entropy of the Voynich MS text was first studied in detail by W.R. Bennett, Yale professor, in the 1970's [10]. The relevance of such an analysis is that entropy values do not change when applying a simple substitution cipher. Thus, if the Voynich MS text were a simple substitution of a plain text in some language, the entropy values of the Voynich MS text should match those of the source language.

The single character entropy may be computed from the frequency distribution of single characters. In Figure 8 we can see the single character distribution of two texts. The first is the Latin text of Mattioli's herbal, a printed book from the 16th century. The characters have been sorted from high to low frequency. The second is a transliteration of a large part of the Voynich MS made by Friedman's First Study Group (FSG).

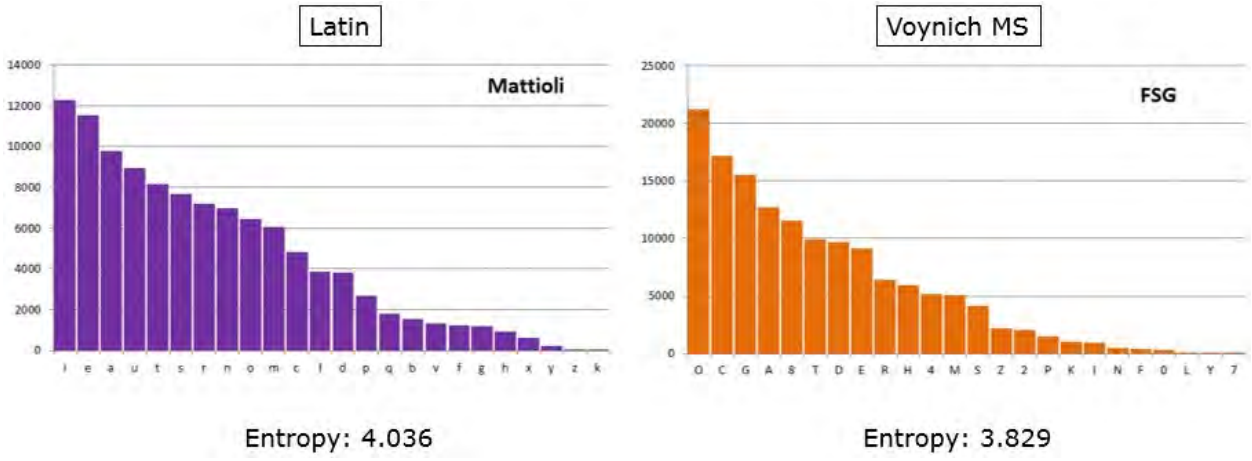


Figure 8: single character distributions of a known Latin text and a Voynich MS transliteration compared

The single character entropy of the Voynich MS text is somewhat lower, but it is not clear how significant this is, and not much can be concluded from this. It becomes more interesting, however, when one looks at the character pair entropy. If subsequent characters in a text appear independently from each other, one can demonstrate that the value of the character pair entropy is twice that of the

[10] See: Bennett, William Ralph: Scientific and Engineering Problem Solving with the Computer. Englewood Cliffs: Prentice-Hall, 1976.

single character entropy. In a meaningful text in any language, this is not the case, and there is some dependency. In that case, the character pair entropy is lower. One may call the difference between the character pair entropy and the single character entropy: the conditional single character entropy. It follows that the value of this quantity will be less than the single character entropy. Table 4 shows the values for three plain texts and again the FSG transliteration of the Voynich MS.

Table 4: unconditional and conditional character entropy values for plain texts and the Voynich MS

Text sample	Entropy	Conditional Entropy
Latin: Mattioli (16th c)	4.036	3.234
Latin: Pliny (classical)	3.998	3.266
Italian: Dante (14th c)	4.014	3.126
Voynich MS (FSG)	3.829	2.052

The character entropy values for the other plain texts are quite similar to that of the Mattioli text. The conditional entropy values for these texts are smaller, and again similar to each other, but in this case the value for the Voynich MS deviates very significantly, and is anomalously low.

So what does this mean? Again, this is just a single number derived from the complete frequency distribution of character pairs, so in order to interpret this significant difference, it will be useful to look at the actual distribution of the character pairs. This may be done visually by representing the frequencies in a square matrix using a colour code for the frequency of each pair. This is shown for the Mattioli text in Figure 9. The first character of each pair is shown vertically from top to bottom and the second character horizontally from left to right. The order of the characters is the same as in Figure 8, i.e. according to decreasing single character frequency. We may concentrate on the square on the left, which shows the actual character pair frequencies, while the square on the right shows what they would have been if the characters would appear randomly in the text, with the given single character frequency distribution. Looking at the left square, one may see for example that the row with the label 'q' only has black squares, except for the column with label 'u', showing that 'q' is always followed by 'u'.

To see more clearly which combinations appear more frequently than 'by chance' and which appear less frequently, one may compute and show the ratio of the values of the two squares. This is shown in Figure 10 below, again for the Latin text of Mattioli. From this figure it appears (as expected) that the high-frequency character pairs tend to be combinations of vowels and consonants.

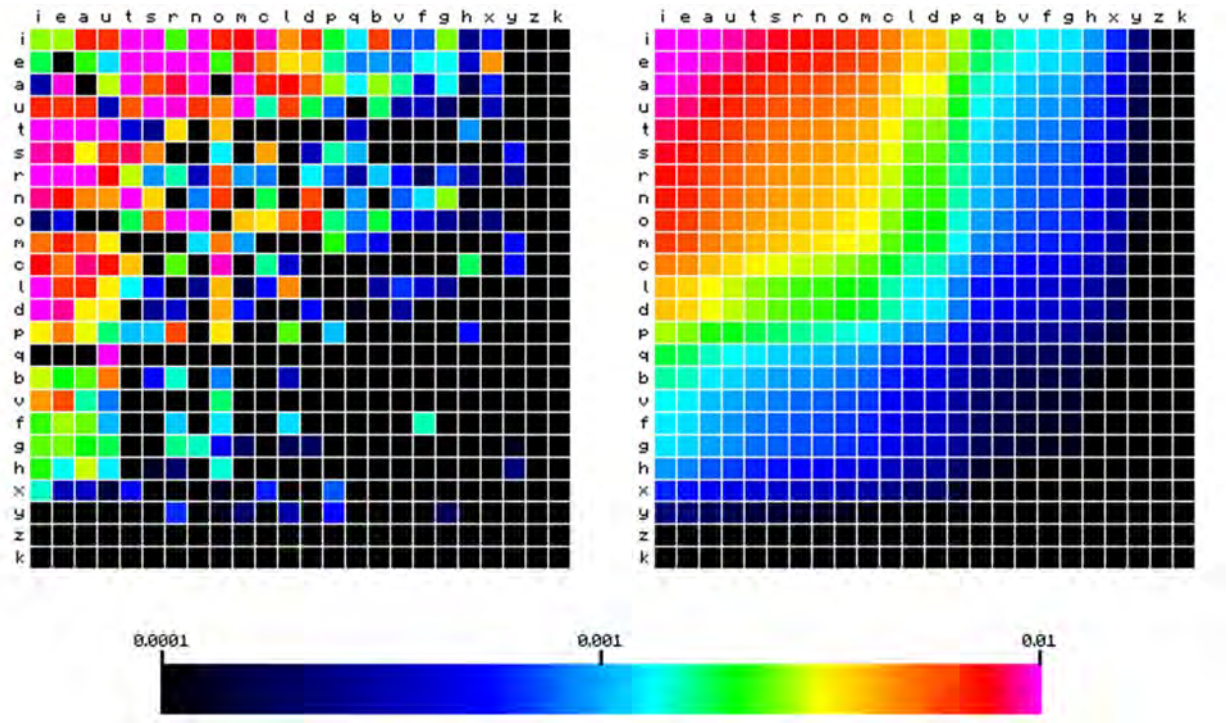


Figure 9: character pair frequency distribution for a Latin text

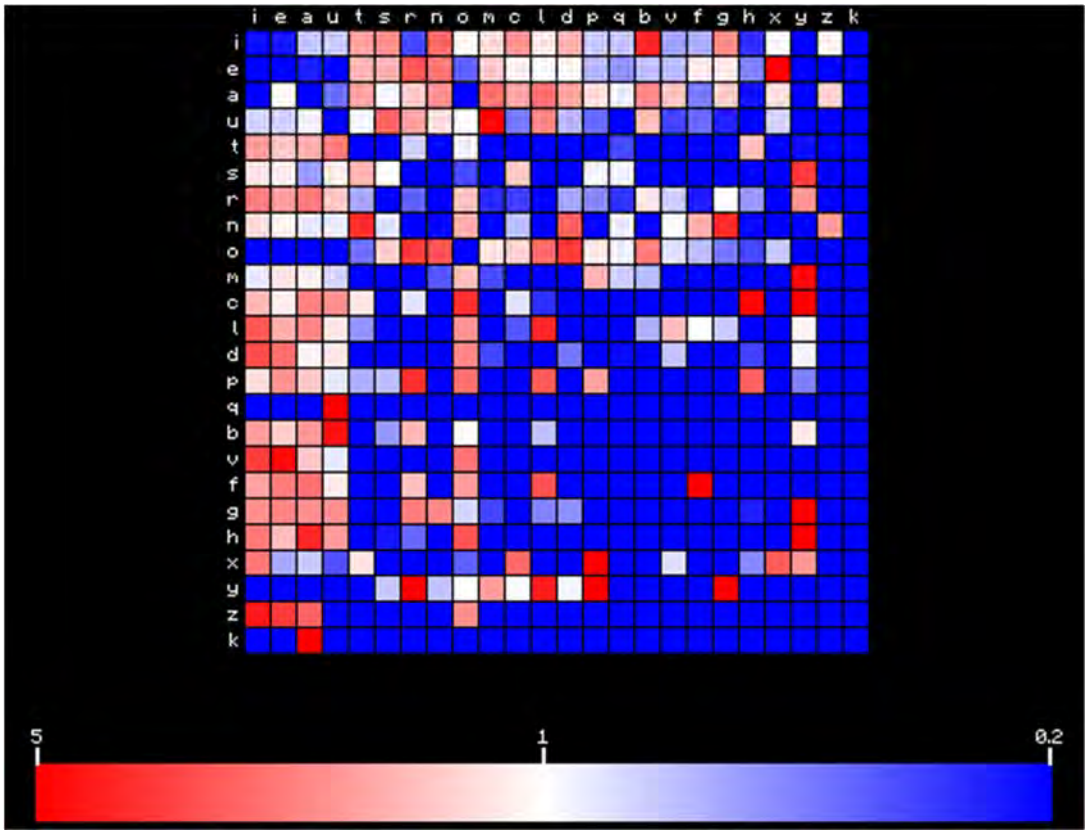


Figure 10: relative character pair frequency for a Latin text

The same figure can be produced for the Voynich MS text. It is show in the right-hand part of Figure 11 below, side by side with the previous figure for the Latin text (on the left).

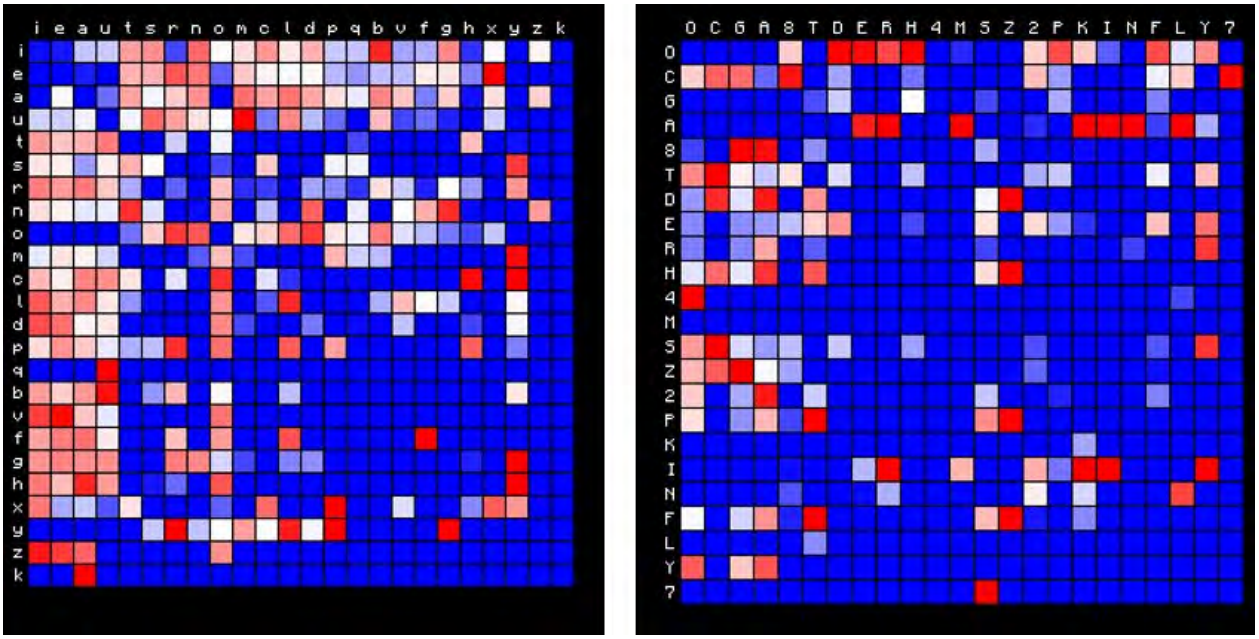


Figure 11: relative character pair frequencies for a Latin text and the Voynich MS compared

This clearly shows a number of very significant differences:

1. The picture for the Latin text is largely symmetric, while this is not at all the case for the Voynich MS text. This shows that the Voynich MS text has a clear preference for the order in which characters can appear in character pairs.
2. The picture for the Voynich MS text is much sparser. There are far fewer valid character combinations.
3. While the Latin picture shows that most consonants can combine with most vowels, there are no clear vowel and consonant blocks in the Voynich MS figure.

The first two points are a reflection of the aforementioned word structure of the Voynich MS text. The third point shows that the common alternation of vowels and consonants in Latin (and other language) texts is not found in the Voynich MS text.

Similar pictures have been made for other plain texts, in Latin, Italian, German and English and they all tend to look like Figure 10. It is again reminded that these figures would not change if the text is subjected to a simple substitution cipher, and it is clear that the Voynich MS text behaves very differently from such plain texts.

Conclusion

On the surface, the text of the Voynich MS looks just like a normal text. However, it behaves very differently:

- There are patterns in the word structure;
- There is no standard alternation between vowels and consonants.

All attempts to decode or translate the text that do not take this into account must necessarily fail.

The innocent nature of the text is its truly remarkable feature.

Acknowledgments

I am grateful to the Beinecke Rare Book and Manuscript Library of Yale University, New Haven CT for being able to study their collections, and for permission to show details of the Voynich MS and associated material (Figures 1 and 4). To the Czech National Library in Prague for permission to use the Kircher letter to Moretus (Figure 3). To the Folger Shakespeare Library in Washington for the interesting workshop and the honour of being invited to participate in it.