The Version of Record of this manuscript has been published and is available in *Cryptologia*; 19 July 2023; DOI:10.1080/01611194.2023.2225716

The Voynich manuscript: discussion of text creation hypotheses

Torsten Timm[a] and Andreas Schinner[b]

[a]Modular Design GmbH, Frankfurt am Main, Germany
[b]Johannes Kepler Universität, Institut für Experimentalphysik, Linz, Austria

ARTICLE HISTORY
Compiled August 2, 2023

ABSTRACT
The *International Conference on the Voynich Manuscript 2022* at the University of Malta has covered many aspects of the current Voynich manuscript research. We discuss a representative selection of the conference proceedings, in particular the question of existing versus non-existing linguistic structures that indicate some underlying meaningful text.

KEYWORDS
Malta Conference 2022; self-citation hypothesis, Voynich manuscript

1. Introduction

The so-called *Voynich manuscript* (VMS) is a handwritten codex, most likely originating in the early Italian Renaissance of the 15th century. It contains drawings of (unidentifiable) plants, astrological or astronomical diagrams, strange arrangements of tubs and pipes, and human figures, mostly nude females. The most striking feature, however, is the elegant unique script that has defied commonly accepted interpretation, decryption or translation so far. For an introduction and survey see, e.g., [7,15,25].

Throughout the last 110 years since its re-discovery by Wilfrid M. Voynich numerous attempts have been made to understand the many puzzling statistical features uncovered by modern computer-aided corpus analysis. Some of them seem to indicate the existence of linguistic structures, while others are very far from what could be expected even for “exotic” languages and/or medieval cryptology. Unfortunately, the two schools of thought, meaningful versus meaningless text, have recently drifted apart to a point, where obviously the desire to find “something legible” in the VMS frequently replaces Occam’s principle by the free adjustment of observations to theories.

To assess the current situation, that Dunin and Schmeh [8] call “a cryptological cold case,” this paper takes a closer look at the snapshot of VMS research presented in the *International Conference on the Voynich Manuscript 2022*, University of Malta. In section 2 we analyze two arguments that are frequently used in favor of the non-meaningless text hypothesis, namely the existence of two distinct “languages” in the VMS, as well as the application of topic modeling techniques.

In section 3 we then discuss eight selected papers of the conference proceedings, which we believe to be representative for the current state of scientific VMS research.

Address correspondence to Torsten Timm. E-mail: torsten.timm@kereti.de

A short conclusions section ends this paper.

Throughout this article the following conventions are used: the term *token* denotes any string of characters (glyphs) separated by spaces or line start or end; a *word* is a type of token regardless of its frequency in the text. Note that in linguistic literature also the synonyms token→*word* and word→*word type* are common. For characters or tokens from the VMS script we use the *European Voynich Alphabet* (EVA) [30]; the letters (or sequences of letters) are printed *italic* between angle brackets. VMS text samples were taken from Takahashi's transcription [26].

2. Linguistic structures

The existence versus non-existence of structures in the VMS that are characteristic for linguistically meaningful text (*and* that cannot be explained as by-product of an algorithmic construction process) is perhaps the most important key question. Several language-related features of Voynichese have previously been proposed as evidence, in particular token frequency or vocabulary buildup scaling laws. Montemurro and Zanette [20] explicitly use Zipf's first law as argument for interpreting the results of their word pattern analysis (an early form of topic modeling) as seemingly supporting the "non-gibberish" hypothesis. Meanwhile it has been shown by [28] that both of Zipf's laws *can* emerge as necessary by-products of an intuitive pseudo-text algorithm. This automatically also rules out Heaps' law, which is a mirror image of Zipf's first law [19], as possible evidence of linguistic structures.

Recently, researchers arguing in favor of the natural language (or encrypted[1] language) text hypothesis very frequently focus on topic modeling [24] and the Currier languages [5]. Indeed, the existence of two statistically strictly separated sub-texts, Currier A and B, *would* provide some evidence for an underlying meaningful text, either as two dialects, topics, or different encryption/encoding schemes. Why should someone with the intention of creating nonsensical pseudo-text invent *two* different methods of doing so?

2.1. The Currier languages

The answer to this question lies in another important one, namely, how well defined the boundaries between Currier A and B domains, respectively, *really* are. The basic principle of topic modeling techniques is the automated, statistically optimized, partitioning of text vocabulary vector space. In the next subsection we will coarsely discuss the potential dangers of applying them as "black box tools" to the VMS. Consequently, we will use a simpler, more transparent method here.

The VMS vocabulary consists of about 7000 words. Let $\vec{v}$ be a vector with each component representing the token frequency of one of these words. Each individual folio uses a subset of the global vocabulary, so most coordinates of the word vector representing a particular folio will be zero. To assess the similarity between two folios we calculate the angle ϕ between the corresponding vectors[2].

$$\cos(\phi) = \frac{\vec{v}_1 \cdot \vec{v}_2}{|\vec{v}_1||\vec{v}_2|} \tag{1}$$

[1]Obviously, in that case the encryption scheme would have to preserve the linguistic structures.

[2]Sometimes this is also called *cosine similarity* or *cosine distance*.

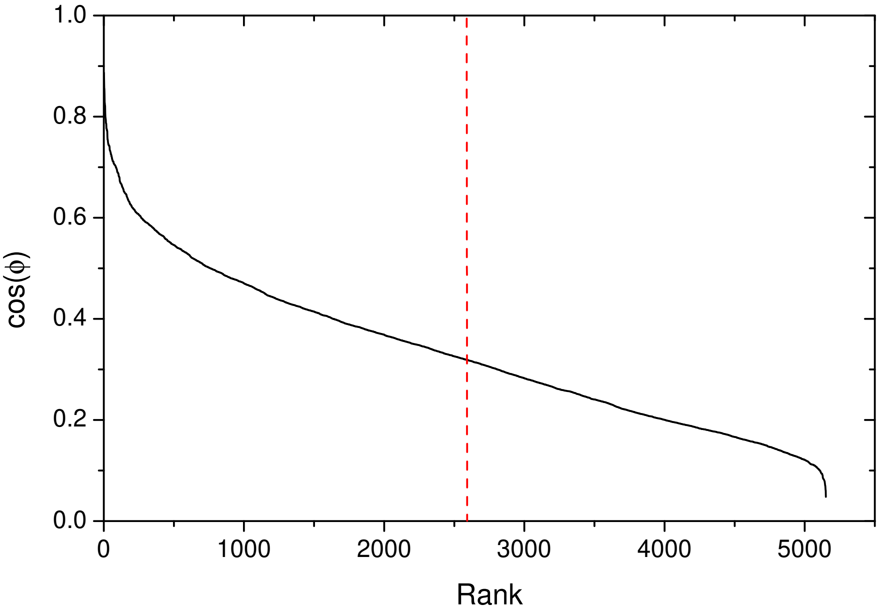

Figure 1. Cosine similarity of all VMS folio pairs versus rank (i.e. sorted to descending order). The vertical red line indicates the hypothetical same/different language pair transition regime; see text.

Since all coordinates are non-negative the resulting cosine will lie between 0 (perpendicular vectors; no word subspace overlap) and 1 (parallel vectors; identical vocabularies and token frequencies). The VMS consists of 102 folios, from which 5151 pairs can be selected (excluding redundancy by symmetry, and the case of two identical folios). Let us, for the moment, assume two well-separated domains, A and B. Then for each pair the two constituent folios can either lie in *the same* or *different* domains. Sorting the $\cos(\phi)$ values to descending order and plotting them versus their rank should then produce a curve that visibly transits from the same-domain pair region (high similarity) to the different-domain regime (low similarity). From the relative percentages of folios attributed as Currier A (53%) and B (39%), respectively, the expected transition regime can easily be located around rank 2600.

However, figure 1 demonstrates the complete absence of any sudden change in the slope of the resulting curve. Instead, besides the lowest and highest ranks (corresponding to exceptionally compatible/incompatible folio pairs) the curve descends smoothly, almost linearly, with increasing rank. This behavior confirms the hypothesis of a continuous evolution from Currier A to B, which had been derived previously [28] on the basis of token/word statistics. Consequently, the Currier languages should be seen as simplified (coarse grained) by-products of the generating algorithm, as well as the scribe's growing experience with it.

Figure 2 shows the actual distribution of the cosine similarity for all VMS folio pairs as a heat map[3]. The black horizontal and vertical lines correspond to missing folios with $\cos(\phi) \equiv 0$. In spite of the coarse resolution some interesting regions can still be identified: The two most self-consistent sections are *Recipes* and *Bio*, with also

[3] A similar plot has been presented by Zandbergen [31], defining the correlation between two pages as the number of tokens common to both pages.

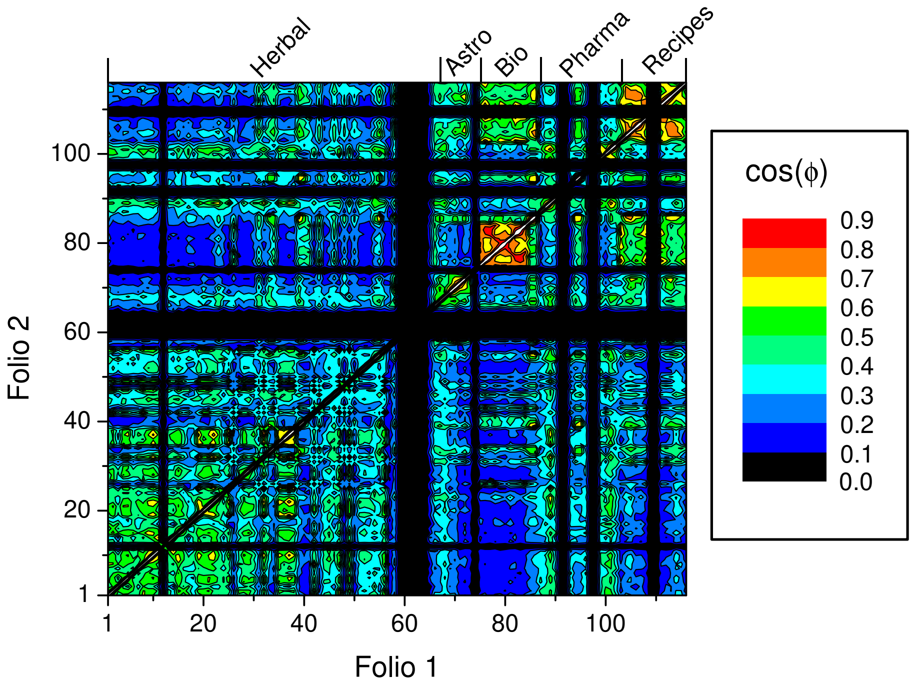

Figure 2. Cosine similarity of all VMS folio pairs.

relatively high similarity between mixed section folio pairs. Of course, this must not be misinterpreted as possible evidence for related textual contents. The answer to the question of whether a particular folio would belong to Currier A is not a definitive *yes* or *no*, but rather a percentage number.

2.2. Topic modeling

Several of the Malta 2022 proceedings implicitly or explicitly refer to a recent application of topic modeling to VMS analysis [24]. Principally, topic modeling is a (much more sophisticated) unsupervised iterative data mining generalization of the rudimentary word space analysis presented in section 2.1. Topic modeling provides an extremely interesting and powerful tool for natural language processing; however, as the term "natural language" already emphasizes, it is very problematic to use the algorithms as "black box oracle" on a corpus potentially containing pseudo language. There are two major dangers: (1) The algorithms are (to a varying degree) mathematically based on the assumption that the texts lie in a statistical range "reasonable" for natural language, and (2) they all depend on an appropriate stop word list.

All topic modeling approaches need a pre-processing step that removes function words from the input data base, because they (a) usually are the most frequent tokens, and (b) carry no contextual information. Too many such words would otherwise bury the rather sensitive clustering algorithm under an intolerable amount of noise, eventually rendering it useless. It is not difficult to define the stop word list for a language you understand, but for Voynichese? The only possibility would be to define an upper token frequency threshold, hoping that the, e.g., 100 most frequent tokens really are function words. However, the token network analysis [28] shows that Voynichese words are connected in a very unusual way, with no graph nodes obviously corresponding to function words[4]. On the other hand, the topic modeling algorithms normally also exclude very low frequency tokens, because they usually are *too* specifically connected to a possible topic to serve as useful anchor points for the respective probability distribution. Note that the VMS text contains a rather high number of hapax legomena, compared to other (natural language) corpora.

Consequently, the results presented in [24] should be understood in the context of these uncertainties. It is especially noteworthy that from the three topic modeling techniques used the *Latent Dirichlet Allocation* (LDA) fails completely (by allocating the entire VMS to a single topic). LDA is the by far most popular approach among the three, most likely because as Bayesian "reverse probabilistic model" it also is the most intuitively constructed one (from the human viewpoint). However, as such it heavily relies on the fact that the investigated texts *really* are natural language. *Latent Semantic Analysis* seems to fare slightly better, but the most pronounced topic categorization can be obtained from applying the very formal (also used in signal processing) method of *Non-negative Matrix Factorization* (NMF) to the VMS.

The three methods all come with known individual strengths and weaknesses, but it is unusual that their results for *one and the same* text differ this widely. It is not surprising (see section 2.1) that the two-topic NMF analysis fails to correctly identify the two Currier language domains, and looking at the fine structures in figure 2 also puts into perspective the reliability of any automated cluster analysis of the VMS. The more unsupervised pattern matching algorithms reach human abilities, the more they become susceptible to pareidolia, too. It is absolutely essential to be able to explicitly check the results of automated statistical processes for plausibility, as has been convincingly demonstrated by Andrew Gelman's famous mock study [12].

3. The conference proceedings

In this section we address a number of selected proceedings of the *International Conference on the Voynich Manuscript 2022*.

3.1. Crux of the MATTR

Lindemann [17] uses two statistical methods, *moving average type-token ratio* (MATTR) and *most common words* (MCW) measure, to assess the morphological complexity of the VMS text (in comparison with natural language samples). His conclusion "The profile of Voynich A suggests that it is more morphologically complex than Voynich B, which may indicate that it encodes a separate language or dialect" (which later on is used as implicit argument against the gibberish interpretation) is based on two assumptions: (1) separable "linguistic domains" Currier A and B, and (2) the position of a text sample in the MATTR/MCW plane characterizes its morphology.

Assumption (1) definitely is wrong, see our analysis in section 2.1, as well as in [28]. As for the validity of (2): the method appears to be rather sensitive to linguistic details of the text, even in the case of *the same* natural language. Table 1 compares two texts written in Latin (that also provides the most natural candidate for a hypothetical VMS clear text message). Both have the same length (10000 tokens), approximately the

[4]See [28]: "No obvious rule can be deduced which words form the top-frequency tokens at a specific location, since a token dominating one page might be rare or missing on the next one."

Table 1. MATTR and MCW values for two Latin samples. From both texts only the first 10000 tokens have been used.

Document	MATTR	MCW	Words	Hapax legomena
Commentarii de Bello Gallico	0.61	0.32	3432	2226
Bibla Vulgata, Genesis	0.47	0.40	2649	1600

same vocabulary size, and even matching hapax legomena/word ratio. Nevertheless, looking at figure 4 of [17] the values in table 1 place Caesar's text in the vicinity of other Latin samples, while the Genesis lies closest to the Iranian language family (and Voynichese).

It is especially problematic that Lindemann argues by interpreting the application of this method to the VMS as if Voynichese had already been proven to be some natural language, while obviously even within samples of the same natural language the results may scatter significantly.

3.2. Voynich paleography

Davis [6] identifies five distinguishable scribal hands, concluding that a group of people had created the VMS. However, there are contradicting viewpoints not discussed by the author, like the observations by Albert Howard Carter from 1946 [3]: "The handwriting is incredibly consistent throughout. It varies only slightly in size, in slant, both of lines and of individual strokes, and in the flow of ink throughout the entire manuscript. It is the work of a single scribe."

Unfortunately, the paper does not disclose many details about which "seventy-four representative pages" are entering the Archetype project (that is not publicly accessible), and only very few analysis examples have been presented. We are aware of the fact that paleography is not an exact science and thus allows for some margin of interpretation, but we find it nevertheless very difficult to verify the author's reported observations. For example, as one characteristic feature to differentiate between two scribal groups (1,3,5 and 2,4) the bowed versus horizontal crossbar of the glyph $<k>$ is discussed. According to Davis a horizontal crossbar indicates that the quill had been lifted after completing the vertical (two strokes), whereas a bowed crossbar more likely is the result of writing the glyph with a single stroke.

However, to our eyes the variability of writing $<k>$ appears very consistent throughout the entire manuscript: For nearly every page it is possible to find instances of $<k>$ written with bowed and with horizontal crossbar, respectively. Additionally, on nearly each page some instances of $<k>$ show an overlap between vertical stroke and crossbar, also indicating two strokes. The paper uses a part of folio f17v to illustrate the handwriting of Scribe 1. The $<k>$-glyphs in figure 3 (a) are from f17v, but they hardly match the description given by Davis: "the [k] character in Scribe 1 is distinguished by a sharp angle at the top of the first vertical as the quill changes direction, a bowed crossbar, a round loop, and a very slight foot at the base of the second vertical."

Later on Davis writes: "the [k] written by Scribe 3 is similar to that of Scribe 1, although slightly more compact." This means that even according to Davis the $<k>$ glyph cannot be used to distinguish between Scribe 1 and 3. Furthermore, part of f105r is used as an example characterizing the writing of Scribe 3. But also there it appears possible to find instances where $<k>$ obviously was formed by two strokes, see figure 3 (b).

With so few research details available, it is impossible to validate or even to assess the results. Consequently, we then find it especially problematic to draw conclusions of

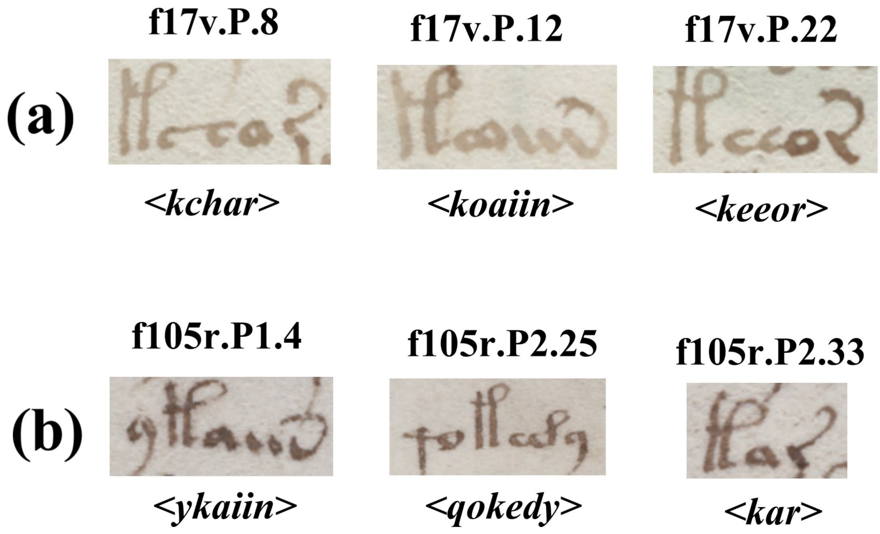

Figure 3. Examples of VMS tokens containing the glyph <*k*>.

high impact, like that "[f] and [p] are abbreviations for [te] and [ke] respectively." Many statistical properties of the VMS indeed are hardly compatible with natural language, but it is dangerous to fit models by introducing additional (auxiliary) hypotheses.

3.3. An analysis of the relationship between words

The paper by Caruana et al. [4] contains a small subset of the much more thorough analysis by Fincher [10], whose results provide strong evidence against the "natural language hypothesis." Already six years earlier Ito's [14] investigation of word pairs has led him to a similar conclusion: "Almost all words have almost same left and right entropy. [...] This result suggests that words in VMS are quite context-independent."

Caruana et al. observe that "the ratio of skewed pairs for the normal document is only slightly more than double that of the random shuffled document," interpreting this result as evidence "that there is some sort of linguistic structure within the manuscript."

Unfortunately, this paper does not discuss the origin of the only seemingly significant asymmetry in the *normal* versus *randomized* "skewed pairs" ratio. In particular, two classes of Voynichese words prefer one sorting order over the other: (1) similar words, e.g., the pair <*shedy*>.<*chedy*> (frequency 8) versus <*chedy*>.<*shedy*> (frequency 3), and (2) two words that also can appear as a single word, e.g. <*olchedy*> (frequency 38): the "natural order" <*ol*>.<*chedy*> (frequency 21) occurs much more often than <*chedy*>.<*ol*> (frequency 3).

Thus only very special words in the VMS prefer a certain order. This becomes even more obvious by investigating repetitive three-word sequences, see [27], page 3. All observations appear perfectly compatible with the hypothesis of an algorithmically

Table 2. Number of necessary token changes for balanced rightwardness.

<*sh*> → <*ch*>		N_Δ	<*qo*> → <*o*>		N_Δ	<*k*> → <*t*>		N_Δ
<*shedy*>	<*chedy*>	40	<*qokaiin*>	<*okaiin*>	9	<*qokeedy*>	<*qoteedy*>	6
<*shey*>	<*chey*>	11	<*qol*>	<*ol*>	23	<*qokedy*>	<*qotedy*>	12
<*shol*>	<*chol*>	21	<*qokeey*>	<*okeey*>	2	<*okaiin*>	<*otaiin*>	22
<*shor*>	<*chor*>	16	<*qokeedy*>	<*okeedy*>	21	<*qokaiin*>	<*qotaiin*>	13

generated pseudo text.

3.4. Grapheme distributions

The statistical analysis by Feaster [9] demonstrates the asymmetric positional distribution of Voynichese tokens with respect to lines and pages. The effect undoubtedly exists and is statistically significant. However, his conclusion "[...] the self-citation hypothesis doesn't account satisfactorily for distribution patterns of the kind described above" is not correct. It is based on the assessment that "[...] repeated arbitrary changes made to strings during copying should have caused the distributions of any variable elements to converge rather than to diverge." This only holds under two assumptions: (1) an infinite copy iteration, and (2) truly random modifications.

According to our theory, both conditions are not fulfilled. The process must necessarily be re-initiated on every new page, and the scribe did certainly not execute the algorithm like a computer. He/she was not only influenced by (most likely even spontaneous) preferences, but also had to react to the text already visible on the partly finished page. Also note that Gaskell and Bowern report [11] "[...] greater biases in character placement within lines and word placement within sections [...]" as the result of an experiment in which students were asked to produce pseudo-text samples.

Table 2 gives a first impression about how significant the observed bias is, with respect to actual token counts. It estimates how many tokens you would have to change in order to nullify the rightwardness imbalance. We have used a rather primitive algorithm here that simply replaces one occurrence of the target token in the leftmost 25% of the line, always starting the search from top of the document. Of course, randomizing the target search, and especially a different positional threshold will change the numbers, but most likely not dramatically. Note that the maximal value 40 still is only about 10% of the frequency of the token <*shedy*>.

It would be trivial to implement the scribe's preference for starting lines with certain variants of the pseudo-words[5] into the simple example algorithm presented in [28][6]. Consequently, the reported distribution pattern is no *experimentum crucis* to decide the key question "gibberish versus non-gibberish."

3.5. Polygraphia III

Essentially, the "Polygraphia" hypothesis proposed by Hermes [13] is based on the assumption that each token of the Voynich manuscript (VMS) encrypts a single letter of an underlying clear text. To decipher the text the Voynichese words then must correctly be arranged in a table with 26 rows (corresponding to the 26 letters of the Latin alphabet) and an unknown number of columns.

[5]Another source for the bias could be afterthought glyph modifications in just finished pages. Note that you can easily change <*ch*> to <*sh*>, and <*k*> to <*t*>, but not vice versa. For a meaningless text such replacements merely would adjust its optical appearance.

[6]It can be done by a simple adjustment of the list of preferred prefixes.

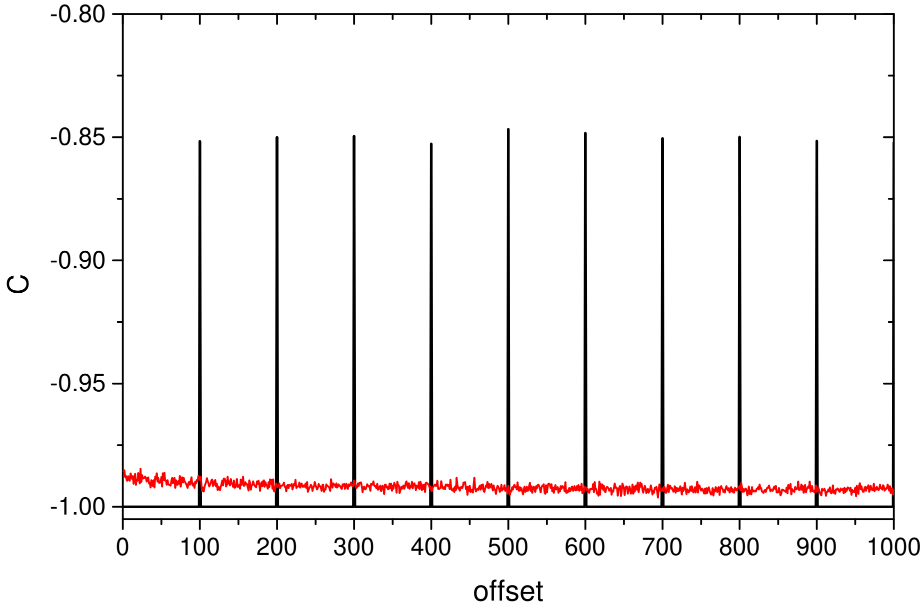

Figure 4. Token autocorrelation versus sequence offset of two documents. Red curve: "Recipes" section of the Voynich manuscript (*f103r-f116v*); black curve: Latin text encrypted by a "Polygraphia III"-like letter → cipher text word mapping (table width = 100 columns).

This hypothesis has several weaknesses, of which the perhaps most critical one is the token autocorrelation problem: Let t_i denote the sequence of tokens in the VMS, then one way (frequently used in cryptanalysis) to define the autocorrelation C of t_i with offset k is

$$C = \frac{A - D}{A + D} \tag{2}$$

where A and D denote the number of agreeing and the number of non-agreeing sequence members, respectively, when comparing t_i with t_{i+k}.

The black line in figure 4 is the result of calculating C for a Latin text, encrypted by a 100 columns table, each cell containing a unique cipher word. Note the distinct peaks at multiples of the table width that are completely absent in the VMS[7] (of course, we have also checked the offset values beyond the limit shown in the figure).

The perfect negative correlation in between the peaks is a consequence of the fact that in this example the encryption is *homographic*[8], i.e., a particular cipher text word always decrypts to the same letter (but not vice versa). However, the more likely case of different columns sharing some entries will merely add (small) random fluctuations to the $C \approx -1$ base line *without* erasing the peaks at multiples of the number of columns. Actually, detecting those peaks is the first step in every automated procedure to crack polyalphabetic ciphers, e.g. of Vigenère type.

---

[7]To avoid discussions about the possibility of different encoding tables corresponding to different sections (Currier languages) we have restricted the calculation to the statistically homogeneous "Recipes" section.

[8]as suggested by table 2 in [13]

So, how can the absence of periodic autocorrelation peaks in the VMS be compatible with any polyalphabetic letter $\rightarrow$ word mapping[9]? Hermes mentions a "random procedure" as basis for the code book table, which then would have to be compatible not only with the absence of autocorrelation effects, but also the preferred position (with respect to the top of the line or paragraph) of some words, as well as the fact that similar words often appear contextual.

Besides the autocorrelation there are additional problems connected to this VMS cipher text hypothesis. For example, to explain even the basic VMS statistics the token distribution would have to match the letter distribution (modified by the table layout) of the clear text language (for which Latin would be the most likely candidate). The VMS text has a relatively high number of hapax legomena. Assuming that they all correspond to low-frequency letters poses significant boundary conditions to possible tables, not to mention the requirements necessary to reproduce all the more advanced statistical properties centered around the complex of token frequency, string similarity, and spatial vicinity.

And finally: the VMS contains about 36000 tokens. If each word represents a clear text letter then said clear text will consist of approximately 4500 tokens (assuming an average Latin word length of approximately 8 letters). To place the about 7000 Voynichese words into a letter encryption table it must have at least 270 columns, most likely significantly more. Such an enormous effort to hide a relatively short message is only conceivable in case of a really important secret. But for what reason would one then "hide" it in such an obvious place? After all, Trithemius' work also had all been about hiding text in inconspicuous prayers or the like.

3.6. Evidence of word structure

VMS word grammars, like that proposed by Zattera [32], are principally problematic: they all tag a significant vocabulary portion of an unknown writing system as "irregular," usually words that appear as concatenation of two "regular" words, or very low frequency words (that usually are seen as errors because they violate the rules just by a single glyph). Several words characteristic for Currier B sections are very rare in Currier A — and would probably be called "irregular" on basis of an A-only word grammar. Furthermore, it is difficult for the word grammar approach to really explain the characteristic relationship between similarity, spatial vicinity, and token frequency. Zattera argues with section-dependent parameters: "One could speculate that the method used to create the text, whether it is an actual cypher or a way to generate gibberish, relies on some parameters e.g., a table, or a grill [...]) which are made different for each section." Consequently, he warns about potential misinterpretation of topic models: "Previous works that used clustering to support the idea of different topics in different sections might be just surfacing a difference which is the result of the mechanism used to generate the word types, not a difference in terminology in the plaintext." (see also section 2.2).

However, we believe our token network approach [28] provides the more appropriate description of the VMS. It offers a "natural" explanation for the similarity/vicinity/frequency correlations as a by-product of the generating copy recursion. Thus the observed token frequencies are consequences of this iterative process, rather

[9] Unless the table has as many columns as there are letters to encrypt. In that case the procedure turns into a *one-time pad* cipher, for which (provided perfect randomness of the key) the impossibility of breaking it can be mathematically shown.

than determined by an external grammar.

3.7. Enciphered after all?

Word-game like enciphering methods that increase the predictability of text have been proposed previously (see, e.g., [23]) throughout the last decades as a possible explanation for the statistical properties of the VMS at the character-level. In spite of the fact that the paper by Bowern and Gaskell [1] apparently follows a similar approach, the methods discussed therein are sometimes denoted as "character-level text metrics," and sometimes also as[10] "unusual word-level predictability." Since the authors also write that words in a linguistic system "combine to form sentences in consistent ways that impart meaning," it remains unclear if the term "word-level predictability" really is always used to refer to character-level metrics, or rather to observations at the word/token-level.

Furthermore, the description of alternative theories [21,22,28] as presented in this paper is strongly misleading: none of the three cited publications use the character-level entropy as an actual argument against a natural text hypothesis: [22] investigates the VMS text on the scale of long-range correlations (covering a characteristic range of at least one text line), as well as string metric similarities at the word-level; Rugg and Taylor [21] propose a generation method based on sets of tables and grilles, while, finally, Timm and Schinner [28] not only present a possible generating algorithm ("self-citation") that successfully reproduces the statistical key properties of the VMS at character- and word-level (and beyond), but also demonstrate how significantly Voynichese deviates from what should be expected for natural language[11].

In addition to this, the authors have cited several publications [2,16,18,24] to confirm their viewpoint that Voynichese would have many properties in common with natural language, as soon as one looks beyond paragraph level. However, Lindemann and Bowern [2,18] never tried to resolve the apparent incompatibility of their proposed linguistic (morphological) structures with the VMS token network (and the significant relation between string and spatial metrics, and token frequency). Layfield [16] only demonstrates that "less probable words contain more informative segments," a feature that our previous paper [28] describes as "high-frequency tokens also tend to have high numbers of similar words," arguing that "this characteristic dependence of token frequency from word similarity is just another manifestation of the long-range correlations that have been uncovered and discussed by several researchers throughout the last decade."

Sterneck et al. [24] actually report a "conjunction of scribe and subject matter" as a conclusion from their cluster analysis. Thus, topic change, together with scribal preferences and "two methods of encoding at least one natural language" (as suggested by Bowern and Lindemann [2]) show the requirement of three different explanations for the change in the VMS token frequencies. This significantly lessens the argumentative strength when these frequencies should serve as evidence for the presence of linguistic structures, in particular that words in the Voynichese text would "form sentences in consistent ways that impart meaning," or that Voynichese should possess "phonology, morphology, semantics, and syntax." This statement had also already been contra-

[10] In [2] the authors write: "Our work argues that the character-level metrics show Voynichese to be unusual, while the word- and line-level metrics show it to be regular natural language and within the range of a number of plausible languages." However, here they explicitly mention "unusual word-level predictability."

[11] "No obvious rule can be deduced which words form the top-frequency tokens at a specific location, since a token dominating one page might be rare or missing on the next one." [28]

dicted by D'Imperio's conclusion from 1978 [7]: "The short words, the many sequential repetitions, the rarity of one- or two-letter words, the rarity of doublets, all militate against simple substitution. So also the strange lack of parallel context surrounding different occurrences of the same word." It is disappointing that not a single one of the aforementioned publications contains even a tentative discussion of the possibility that in reality the token frequencies might undergo a continuous evolution from Currier language A to B, rather than being the consequence of two distinct encoding schemes and/or scribes (see section 2.1 and [29]).

Essentially, the paper discussed here merely studies various word-game like character-based substitutions (like adding null characters, removing vowels, remapping prefixes/suffixes, etc.) to demonstrate that such a transformation indeed will increase the character-level predictability. No arguments are presented that any of the described methods might actually have been used in the VMS text[12]. Such a result is hardly surprising. But what does it say about the VMS?

3.8. Gibberish after all?

It is unusual that the same two authors submit two papers [1,11] to the same conference proceedings that are contradicting one another this much. While the outcome of the experiment[13] described in [11] strengthens our viewpoint, it is nevertheless necessary to clarify a few points.

Gaskell and Bowern write: "Informal interviews and class discussions confirmed that many participants did indeed adopt this type of approach[14] to create their texts, although they generally did so intuitively rather than by developing an explicit algorithm such as that published by Timm and Schinner." However, we also assume that the author of the VMS had developed the algorithm intuitively, rather than on a conceptual basis, with growing experience amplifying the peculiarities of the copying process. Unfortunately, the paper does not report the test persons' motivation for the word repetitions. In [2] Bowern and Lindemann hint: "We tested this point in an undergraduate class and found that beyond about 100 words, the task of writing language-like non-language is very difficult. It is too easy to make local repetitions [...]" This is an important point, because it clarifies that any scribe creating language-mimicking gibberish will sooner or later replace the tedious task of inventing more and more words by the much easier reduplication of existing text (and stick with this strategy).

Nevertheless, later on in their paper Gaskell and Bowern state: "A more significant limitation of this work is that, because of the short length of our text samples, we are unable to test whether gibberish can replicate the larger structural features (such as "topic words") which have been observed in the VMS. At present, these features pose a serious challenge to proponents of the hoax hypothesis." While they do not explicitly explain (or give examples) for the term "topic words" in this context, we presume that it refers to "topic modeling", or previous attempts to associate some Voynichese words with particular illustrations and/or Currier languages. We have already discussed the problems connected with this approach in section 2. Neither is it possible to identify commonly used words that *exclusively* appear in sections labeled as Currier A or B,

[12] Actually the paper does not refer to any word of the VMS text. It only lists some Z-score metrics and states that the "individual metrics vary substantially."

[13] Volunteers were recruited to write short "gibberish" documents, as a basis for a statistical comparison with the VMS and linguistically meaningful texts.

[14] This refers to the "self-citation" algorithm [28].

respectively, nor words that can be found on pages with certain illustrations only.

4. Conclusions

It is impossible to mathematically prove that a particular set of strings is truly meaningless, because this would need a general method to compute an upper boundary to the Kolmogorov complexity. Consequently, the only way to disprove the meaningful (encrypted) text hypothesis is by exclusion of alternatives. In the case of the VMS many statistical correlations are so strange (with respect to natural language or even simple word based encryption schemes) that the interpretation as meaningless text should indeed appear as the most promising one. Even more so as, to our knowledge, not a single unimpeachable piece of evidence against it has been found so far.

Nevertheless, many researchers obviously *wish* to discover some legible information in the VMS, and they have difficulty even considering another possibility. There exists a definite red line between scientific research and belief systems that should never be crossed. It is a bit discouraging how many conference contributions have followed argumentative chains with obviously weak or even clearly broken links, just to pursue a preferred theory.

From our viewpoint, the most likely origin of the VMS text is still some kind of "self reproduction process." The algorithm outlined in [28] does not reproduce some of the more subtle statistical observations (as several of our critics have emphasized), but this does not disprove the underlying concept. Perhaps we should develop an updated version with additional statistical details implemented, but even then there will always be another new deviating fine structure or local aberration. The VMS was created by a human being with all freedom to spontaneously make and break (intuitive or explicit) own rules. After all, from the viewpoint of complexity theory, most likely the minimal representation of the VMS is the manuscript itself.

About the Authors

Torsten Timm, diploma in computer science at the Friedrich-Alexander University of Erlangen-Nürnberg, works as a computer scientist in Dresden, Germany. In 2005 he published a book about his research on the Phaistos Disc. His previous research on the Voynich Manuscript focused on a detailed analysis of co-occurrence patterns for similar tokens. In 2014 he published a new generation method for the text in the Voynich Manuscript.

Dr. Andreas Schinner is a theoretical physicist, performing freelance research at the Johannes Kepler University in Linz, Austria. His main area of scientific interest is theoretical solid state physics — particularly particle beam interactions with matter. He is also working as a self-employed software developer.

References

[1] Bowern, C. L. and D. E. Gaskell, 2022. "Enciphered after all? Word-level text metrics are compatible with some types of encipherment." *International Conference on the Voynich Manuscript 2022*, University of Malta. https://ceur-ws.org/Vol-3313/paper6.pdf (accessed February 1, 2023).

[2] C. L. Bowern and L. Lindemann, 2021. "The Linguistics of the Voynich Manuscript," Annu. Rev. Linguist., vol. 7, no. 1, pp. 285–308. doi:10.1146/annurev-linguistics-011619-030613

[3] Carter, A. H, 1946. "Some impressions of the Voynich Manuscript." https://voynich.net/reeds/docs/carter.txt (accessed January 6, 2023); presently only accessible via https://web.archive.org/.

[4] Caruana, A., C. Layfield, and J. Abela, 2022. "An Analysis of the Relationship between Words within the Voynich Manuscript." *International Conference on the Voynich Manuscript 2022*, University of Malta. https://ceur-ws.org/Vol-3313/paper8.pdf (accessed February 1, 2023).

[5] Currier, P. H., 1976. "New Research on the Voynich Manuscript." In *Proceedings of a Seminar held on 30 November 1976*, ed. M. E. D'Imperio. Washington, DC. Privately Printed Pamphlet. November 30, 1976.

[6] Davis, L. D., 2022. "Voynich Paleography." *International Conference on the Voynich Manuscript 2022*, University of Malta. https://ceur-ws.org/Vol-3313/keynote2.pdf (accessed February 1, 2023).

[7] D'Imperio, M. E. 1978. *The Voynich Manuscript – An Elegant Enigma.* Laguna Hills, CA: Aegean Park Press.

[8] Dunin, E. and K. Schmeh, 2022. "The Voynich Manuscript Compared with Other Encrypted Books." *International Conference on the Voynich Manuscript 2022*, University of Malta. https://ceur-ws.org/Vol-3313/paper15.pdf (accessed February 1, 2023).

[9] Feaster, P., 2022. "Rightward and Downward Grapheme Distributions in the Voynich Manuscript." *International Conference on the Voynich Manuscript 2022*, University of Malta. https://ceur-ws.org/Vol-3313/paper12.pdf (accessed February 1, 2023).

[10] Fincher, M., 2008. "Word Pair Permutation Analysis of natural language samples and its value for characterizing the 'Voynich Manuscript.' " https://studylib.net/doc/5850543/word-pair-permutation-analysis-of-natural-language-sample... (accessed February 1, 2023).

[11] Gaskell, D. E. and C. L. Bowern, 2022. "Gibberish after all? Voynichese is statistically similar to human-produced samples of meaningless text." *International Conference on the Voynich Manuscript 2022*, University of Malta. https://ceur-ws.org/Vol-3313/paper4.pdf (accessed February 1, 2023).

[12] Gelman A. and G. A. Romero, 2010. " 'How many zombies do you know?' Using indirect survey methods to measure alien attacks and outbreaks of the undead." arXiv:1003.6087 [physics.soc-ph]. doi:10.48550/arXiv.1003.6087

[13] Hermes, J., 2022. "Polygraphia III: The cipher that pretends to be an artificial language." *International Conference on the Voynich Manuscript 2022*, University of Malta. https://ceur-ws.org/Vol-3313/paper7.pdf (accessed February 1, 2023).

[14] Ito, A., 2002. "Observation of left and right entropy in Voynich MS." https://www.ic.unicamp.br/~stolfi/links/programs/c/NEUROMAT/projects/voynich/paper021209.pdf (accessed February 1, 2023).

[15] Kennedy, G. and R. Churchill 2005. *The Voynich Manuscript: The unsolved riddle of an extraordinary book which has defied interpretation for centuries.* London: Orion Publishing Group Ltd.

[16] C. Layfield, L. van der Plas, M. Rosner, and J. Abela, 2020. "Word Probability Findings in the Voynich Manuscript." In Proceedings of LT4HALA 2020 – 1st Workshop on Language Technologies for Historical and Ancient Languages, Marseille, France, pp. 74–78. https://www.aclweb.org/anthology/2020.lt4hala-1.11 (accessed February 1, 2023).

[17] Lindemann, L., 2022. “Crux of the MATTR: Voynichese Morphological Complexity.” *International Conference on the Voynich Manuscript 2022*, University of Malta. `https://ceur-ws.org/Vol-3313/paper9.pdf` (accessed February 1, 2023).

[18] L. Lindemann and C. L. Bowern, 2021. “Character Entropy in Modern and Historical Texts: Comparison Metrics for an Undeciphered Manuscript.” arXiv, May 18, 2021. doi: 10.48550/arXiv.2010.14697.

[19] Lü, L., Zi-Ke Zhang, and T. Zhou, 2010. “Zipf’s Law leads to Heaps’ law: Analyzing Their Relation in Finite-Size systems.” *PLoS ONE* 5 (12): e14139. doi:10.1371/journal.pone.0014139

[20] Montemurro, M. A., and D. H. Zanette, 2013. “Keywords and co-occurrence pPatterns in the Voynich Manuscript: An information-theoretic analysis.” *PLoS ONE* 8 (6):e66344. doi:10.1371/journal.pone.0066344.

[21] Rugg, G., and G. Taylor 2017. “Hoaxing statistical features of the Voynich Manuscript.” *Cryptologia*, 41 (3):247–68. doi:10.1080/01611194.2016.1206753.

[22] Schinner, A., 2007. “The Voynich Manuscript: Evidence of the Hoax Hypothesis.” *Cryptologia*, 31 (2):95–107. doi:10.1080/01611190601133539.

[23] Stallings, D. J., 1997. EKT hypothesis. `http://ixoloxi.com/voynich/ekt.txt` (accessed February 1, 2023).

[24] Sterneck, R., A. Polish, and C. Bowern, 2021. “Topic Modeling in the Voynich Manuscript.” arXiv:2107.02858 [cs.CL]. doi: 10.48550/arXiv.2107.02858.

[25] Stolfi, J., 2003. “Voynich manuscript Stuff.” `http://www.ic.unicamp.br/~stolfi/voynich/` (accessed February 1, 2023).

[26] Takahashi, T. VMS-Transcription. `http://voynich.freie-literatur.de/index.php?show=extractor`. (accessed February 1, 2023).

[27] Timm, T. 2014. “How the Voynich Manuscript was created.” arXiv:1407.6639 [cs.CR]. doi:10.48550/arXiv.1407.6639.

[28] Timm, T. and A. Schinner, 2020. “A possible generating algorithm of the Voynich manuscript.” *Cryptologia* 44 (1): 1-19.

[29] Timm, T. and A. Schinner, 2021. “Review of *The linguistics of the Voynich manuscript* by Claire Bowern and Luke Lindemann.” *Cryptologia* 45 (5): 434-438.

[30] Zandbergen, R., 2003. “The Voynich manuscript.” `http://www.voynich.nu` (accessed February 1, 2023).

[31] Zandbergen, R., 2003. “Currier A and B: two different languages?” `https://www.voynich.nu/extra/lang.html` (accessed February 1, 2023).

[32] Zattera, M., 2022. “A New Transliteration Alphabet Brings New Evidence of Word Structure and Multiple ’Languages’ in the Voynich Manuscript.” *International Conference on the Voynich Manuscript 2022*, University of Malta. `https://ceur-ws.org/Vol-3313/paper10.pdf` (accessed February 1, 2023).