

International Conference on the Voynich Manuscript 2022, Abstracts



Keynote

- René Zandbergen: Transliteration of the Voynich MS text

Session One

- Alexander Boxer: Fingerprinting Gibberish: A Quantitative Comparison of the Voynich and Sloane MS 3188.
- Koen Gheuens and Cary Rapaport: Above and Beyond Voynich Canopies: Tents as a Recurring Motif in Beinecke MS 408.
- Keagan Brewer: 'I beg your grace to suppress this chapter or else to have it written in secret letters': The emotions of encipherment in late-medieval gynaecology.
- Daniel Gaskell and Claire Bower: Gibberish after all? Voynichese is statistically similar to human-produced samples of meaningless text.

Session Two

- Kevin Farrugia, Colin Layfield and Lonneke van der Plas: Demystifying the scribes behind the Voynich Manuscript using Computational Linguistic Techniques.
- Claire Bower and Daniel Gaskell: Enciphered after all? Word-level text metrics are compatible with some types of encipherment.
- Jürgen Hermes: Polygraphia III: The cipher that pretends to be an artificial language.
- Andrew Caruana, Colin Layfield and John Abela: An Analysis of the Relationship between Words within the Voynich Manuscript.

Session Three

- Luke Lindemann: Crux of the MATTR: Voynichese Morphological Complexity.
- Massimiliano Zattera: A new transliteration alphabet brings new evidence of word structure and multiple "languages" in the Voynich manuscript.
- Katie Painter and Claire Bower: Examining the history of Voynich glyphs using phylogenetic methods.
- Patrick Feaster: Rightward and Downward Grapheme Distributions in the Voynich Manuscript.

Session Four

- Tavi Stafford: Seven Habits of Highly Eccentric Paragraphs.
- Farley Katz: From Voynich to the Beinecke, the Trail of Ownership.
- Klaus Schmeh and Elonka Dunin: The Voynich Manuscript Compared with Other Encrypted Books.
- Stefan Guzy: Book transactions of Emperor Rudolf II 1576-1612. New findings on the earliest ownership of the Voynich manuscript.

Keynote

- Lisa Fagin Davis: Voynich Paleography.

Keynote – 30 November 2022, 1330-1415 CET

René Zandbergen

Transliteration of the Voynich MS text.

Abstract:

The mysterious writing in the Voynich MS has resisted all attempts to translation or decryption, in spite of the increasing number of attempts over the last decades. What has become clear is that the text has a number of unusual properties, and increasingly advanced text and language analysis techniques have been deployed in order to get to the bottom of this mystery. The basis for all these attempts is the existence of a computer-readable version of the text – a transcription, or rather a transliteration as we don't even know the alphabet of this writing. This paper deals with a number of aspects of the transliteration of the Voynich MS text.

First, a brief historic overview introduces the various efforts and the existing transliterations resulting from them. This also serves to introduce the specific problems that exist with this text. The transliterations use different conventions for the representation of the text (usually called 'alphabets'), but their accuracy is largely unknown, and the assumptions that were made in creating these transliterations were likely sub-optimal or even incorrect. This limits or even biases all computer analyses based on them. This is also true for the most popular alphabet: "Eva", which was defined by Gabriel Landini and the author of this paper, just before the year 2000.

Nowadays, there is a multitude of users of these transliteration files, and these users are often unaware of these risks. Many prefer to use whatever transliteration file is most easily available or most easily usable. More accurate transliterations exist, but these are often considered overly complicated to use.

As a next step, this paper introduces a superset of all existing transliteration alphabets, which allows all transliterations to be represented in the same 'super alphabet'. As a result, the completeness and consistency of all existing transliterations can be verified in detail. Some key statistics about the accuracy and completeness of all existing transliterations are presented.

This 'super alphabet' also allows an easy transformation from any existing transliteration into any of the existing alphabets. Furthermore, it allows a user to easily create his own character set definition and convert any existing transliteration into it. With this, it is now possible to repeat computerised attacks on the text quickly and easily for any number of 'alphabets', allowing the analyst to tune this alphabet as he progresses. An example of such an analysis is presented.

Session One – 30 November 2022, 1415-1445 CET

Alexander Boxer

Fingerprinting Gibberish: A Quantitative Comparison of the Voynich and Sloane MS 3188.

Abstract:

Texts that resemble natural language but convey no semantic content — that is, gibberish — still exhibit statistical fingerprints allowing them to be identified from other, superficially similar texts. This paper summarizes the results of a quantitative comparison between the Voynich and the extended passages of “Enochian” transcribed by John Dee and Edward Kelley in Sloane MS 3188. Preliminary to this analysis, and by far its principal labor, is a new transcription of the Voynich text which I have prepared and made freely available on GitHub. Using this transcription, and a transcription of the relevant sections of Sloane MS 3188, several important similarities can be observed. Most notably, both the Voynich and Enochian exhibit an unusual degree of epizeuxis, or token repetition, including triplets. However, the running fraction of hapax legomena, a key statistical fingerprint which distinguishes Enochian from natural language texts, also distinguishes Enochian from the Voynich. Our analysis therefore suggests that even if the Voynich is gibberish, it does not belong the same family of gibberish as Enochian.

Session One – 30 November 2022, 1445-1515 CET

Koen Gheuens and Cary Rapaport

Above and Beyond Voynich Canopies: Tents as a Recurring Motif in Beinecke MS 408.

Abstract:

This paper aims to improve our understanding of a recurring motif in the Voynich Manuscript: canopies. We will select two groups of canopy-like structures from different sections of the manuscript, which have significant overlap between their features. We will analyze these examples in comparison to medieval imagery of tent canopies, demonstrating that 15th century tent construction, decoration, and potentially symbolism, likely inspired elements of Voynich structures and diagrams. Understanding this connection to medieval visual and material culture will grant a deeper understanding of the thoughts behind this enigmatic imagery, and further anchor the Voynich within its historical context.

Session One – 30 November 2022, 1515-1545 CET

Keagan Brewer

'I beg your grace to suppress this chapter or else to have it written in secret letters': The emotions of encipherment in late-medieval gynaecology.

Abstract:

The Voynich illustrations feature hundreds of naked women, some of whom have objects adjacent to or unambiguously pointed towards their genitalia. Taking its prompt from these illustrations, this paper investigates the obscurantist emotions evident in self-censorship, erasure, and encipherment in gynaecological and sexological texts and manuscripts. These subjects were often referred to as 'women's secrets', particularly in Germanic cultural contexts, where the Voynich manuscript may have originated. Examples of encipherment, erasure, and self-censorship in gynaecological or sexological texts will be explored. Dr Johannes Hartlieb (c. 1410–1468), for instance, consistently obscured matters relating to coitus, women, and plants throughout his oeuvre, and called for encipherment of methods for contraception, abortion, and sterilisation. Luke Demaitre observed that some Germanic authors considered women's sexual anatomy alluring but dangerous. It is possible that the Voynich authors were motivated by similar emotions relating to women's secrets.

Session One – 30 November 2022, 1545-1615 CET

Daniel Gaskell and Claire Bower

Gibberish after all? Voynichese is statistically similar to human-produced samples of meaningless text.

Abstract:

The text of the Voynich Manuscript (VMS) has often been regarded as too non-random to be meaningless. However, if the VMS is indeed a hoax, it was probably not produced by a purely random process but rather by some form of automatic writing or glyptolalia in which the scribe(s) simply invented meaningless text as they went based on an intuitive impression of what written language ought to look like. Here, we show that such intuitive “gibberish” is significantly non-random and in fact exhibits many of the same statistical peculiarities as Voynichese. We recruited 42 volunteers to write short “gibberish” documents and statistically compared them to several transcriptions of the VMS and a large corpus of linguistically meaningful texts. We find that “gibberish” writing varies widely in its statistical properties and, depending on the sample, is able to replicate either natural language or Voynichese across nearly all of the metrics which we tested, including traditional criteria for identifying natural language such as Zipf’s law. However, gibberish tends to exhibit lower total information content than meaningful text; higher repetition of words and characters, including triple repeats; greater biases in character placement within lines and word placement within sections; positive autocorrelation of word lengths (i.e., a tendency for words to cluster short-short-longlong rather than short-long-short-long); and a weaker average fit to Zipf’s law. The majority of these properties are also observed in Voynichese. A machine-learning model trained to distinguish meaningful text from gibberish in our dataset identified most VMS transcriptions as more closely resembling gibberish than meaningful text. We argue that these results refute the idea that the low-level linguistic structure of the VMS text is too non-random to be meaningless. However, our writing samples are too short to test whether the higher-level structure of VMS pages and quires could also be produced by gibberish.

Session Two – 30 November 2022, 1615-1645 CET

Kevin Farrugia, Colin Layfield and Lonneke van der Plas

Demystifying the scribes behind the Voynich Manuscript using Computational Linguistic Techniques.

Abstract:

Earlier work studies the paleography of the Voynich manuscript and proposes five scribes. In this work, we use computational linguistic models that are based on the character sequences found on the pages, inspired by work on automatic authorship attribution, as opposed to the earlier work focusing on the nature of the writing script. Our rationale is that if these two independent methods should lead to similar results, we have strong evidence for the earlier identification of the scribes in the Voynich manuscript. We carry out a machine learning experiment where the manuscript is divided into an equal number of pages for three of the five scribes. Four machine learning classifiers create models that have been trained on the classification produced in earlier work to predict a scribe for each page using character n-grams as features. The results of each classifier are analysed comparatively with one another and compared with the classification provided based on paleographic work. The results show when the classifiers produce the same classification as earlier paleographic work on a held-out dataset.

Session Two – 30 November 2022, 1645-1715 CET

Claire Bown and Daniel Gaskell

Enciphered after all? Word-level text metrics are compatible with some types of encipherment.

Abstract:

Many Voynich manuscript analyses have relied on statistical properties of the text to distinguish enciphered natural language from non-language (or gibberish). Schinner (2007) and Rugg & Taylor (2016) have argued that Voynichese is unlikely to be natural language because of its extreme predictability. Conversely Bown & Lindemann (2021), Sterneck et al. (2021), Layfield (2021) and others focus on topic modeling and larger textual units, showing that beyond the paragraph level, Voynichese has many properties in common with enciphered natural language. The question then becomes whether one can discover ciphers that produce the textual characteristics that make Voynichese unusual at the word level, while preserving topic structure across a larger sample. To this end, we investigate the statistical properties of 22 methods of textual manipulation on a sample of historical and contemporary texts. For consistency of comparison we use the same metrics as . While many historical encipherment methods (such as substitution ciphers) are phonological structurepreserving (and therefore not tested here), others, such as the Crema cipher, are not. Results show that there are multiple types of encipherment which reduce conditional entropy; the encoding of multiple phonemes (or orthographic characters) as bigraphs, for example, lowers character entropy to the levels seen in Voynichese, for Latin-encoded texts. Adding null characters (in some patterns) also increases predictability of word formation. While such results do not “prove” that the Voynich manuscript is enciphered, it indicates that the unusual word-level predictability highlighted in previous work is not conclusive evidence that the Voynich manuscript is gibberish.

Session Two – 30 November 2022, 1715-1745 CET

Jürgen Hermes

Polygraphia III: The cipher that pretends to be an artificial language.

Abstract:

This paper demonstrates the existence of a cipher method in the early modern period (Polygraphia III by Johannes Trithemius), which – applied as a random procedure – is able to produce a text that can mimic the oblique properties of the so-called Voynich Manuscript (VMS). This result is quite exciting since it brings back into play highly-debated approaches claiming the existence of hidden comprehensible information within the text of the VMS (which is often referred to as Voynichese). The paper briefly outlines some of the most salient and difficult-to-explain statistical properties of Voynichese, shows how Trithemius stepwise developed a cipher system whose application looks like an artificial language, points out how an application of this cipher generates a text that comes very close to the statistical properties of Voynichese and finally discusses possible starting points of cryptanalytic attacks on a cipher that operates similar to the Polygraphia III encryption.

Session Two – 30 November 2022, 1745-1815 CET

Andrew Caruana, Colin Layfield and John Abela

An Analysis of the Relationship between Words within the Voynich Manuscript.

Abstract:

This paper investigates the presence of linguistic structure within the Voynich Manuscript by analysing various properties of word-pairs found in the manuscript as well as in other works written in natural languages. Apart from the Voynich, the other manuscripts we analysed are the Bible, Dante's *La Divina Commedia*, and Shakespeare's *Macbeth* and *Julius Caesar*. The order of words in the word-pairs in these manuscripts is analyzed and this analysis indicated that several pairs were more likely to appear in one order than the other. These are called 'skewed pairs'. The ratio of the number of skewed pairs in each work to the number of all pairs in each work is plotted, along with this same ratio but for random shuffles of each work. The results indicated that there is a substantial difference in all natural language documents between their normal and shuffled counterparts. The difference is not as large within the Voynich Manuscript but the word-pair occurrence ratio of the original is still considerably higher than the ratio of the shuffled manuscript. This could indicate that the Voynich Manuscript is not random text but may be a language or a cipher.

Session Three – 1 December 2022, 1300-1330 CET

Luke Lindemann

Crux of the MATTR: Voynichese Morphological Complexity.

Abstract:

This study examines Voynichese at the word level, using a corpus of language samples and word-level statistics in order to identify the most plausible language families for Voynichese, and to exclude statistically improbable language families. The results narrow down the possibilities for a source language and emphasize the likelihood that Voynichese does in fact encode meaningful language. Comparison texts include samples from 160 modern languages in ten major language families, as well as historical manuscripts written in Hebrew, Italian, Old Church Slavonic, and Welsh. The methodology employs two particular word distribution statistics, the Moving Average Type-Token Ratio (MATTR) and the Most Common Words percentage (MCW). A graph of the corpus language samples shows a clustering by language family and correlates with the presence of agglutinating or isolating morphology in a given language family. The methodology ranks plausible language families and excludes statistically improbable language families based on their distribution, and indicates that Voynichese possesses a middle range of morphological complexity that most closely resembles that of medieval Germanic languages. By contrast, Semitic languages like Hebrew and Arabic, and languages from the Slavic and Celtic families are possible but less likely candidates. Families that feature heavy agglutinating morphology like Turkic and Uralic can be effectively excluded.

Session Three – 1 December 2022, 1330-1400 CET

Massimiliano Zattera

A new transliteration alphabet brings new evidence of word structure and multiple "languages" in the Voynich manuscript.

Abstract:

The question of which glyphs are actual single Voynich characters is still very open; its correct answer will greatly impact statistical analysis of the text and will guide deciphering attempts. This research uses a computational approach, including machine learning techniques, to create a new transliteration alphabet. It is shown how the structure of Voynich words can be described by assuming each word type is composed of 12 "slots", each being empty or containing specific glyphs. 86.6% of Voynich tokens exhibit this structure and only 1.5% of tokens are words not following this pattern and appearing at least twice in the text. It is therefore assumed that glyphs appearing in slots are the characters of the Voynich alphabet and as such have been mapped into single characters of a newly created transliteration alphabet. This alphabet has been used by an optimization algorithm to automatically produce a formal grammar for words in the Voynich. This grammar is the one with the highest F1 score among those surveyed. Separate grammars have been generated for different sections of the manuscript, and their differences used in a decision tree algorithm which is able to correctly classify pages into their section by looking at the occurrence of four short char sequences. This shows that the difference between sections is not a difference in Voynich vocabulary, as past research based on clustering suggested, rather the vocabularies are different because of differences in inner structure of words. This is evidence for the existence of different "languages" in the manuscript, of which Currier's A and B are only major divisions. The implications of these findings are then discussed.

Session Three – 1 December 2022, 1400-1430 CET

Katie Painter and Claire Bower

Examining the history of Voynich glyphs using phylogenetic methods.

Abstract:

Paleographic analytical methods are used in dating manuscripts and discovering information about their geographic origins. The evolution of script traditions is well understood; for example, the features that separate Uncial from Carolingian, Gothic, Beneventan and Humanistic hands (among other traditions) include the shape of script glyphs, the use of serifs and ligatures, and the placement, use, and size of descenders. In this paper, we use phylogenetic network methods to identify manuscript clusters, based on features of the scripts, and compare them with glyphs used in the Voynich manuscript. Results indicate that the Voynich glyphs do not clearly cluster within any single tradition.

Session Three – 1 December 2022, 1430-1500 CET

Patrick Feaster

Rightward and Downward Grapheme Distributions in the Voynich Manuscript.

Abstract:

A systematic approach is outlined for detecting patterns of uneven positional distribution of words and glyphs within lines and paragraphs in terms of “rightwardness” (distance towards the right end of a line) and “downwardness” (distance towards the bottom of a paragraph). Three case studies are developed to support an argument that such patterns are both more numerous and more pervasive in the Voynich Manuscript than generally supposed and that the specific combination of glyphs that make up a Voynichese word is significantly constrained by that word’s position rightward and downward.

Session Four – 1 December 2022, 1500-1530 CET

Tavi Stafford

Seven Habits of Highly Eccentric Paragraphs.

Abstract:

This paper investigates the unexpected patterns in how the Voynich scribes use “gallows glyphs”. It explores how the position of a word in a paragraph affects the behaviour of these glyphs, and how this varies depending on which gallows character and which scribe is involved. Notable eccentric habits are: excess reliance on gallows glyphs as paragraph initials; high inconsistency with initials elsewhere; preference for the top row of paragraphs; correlation between position in the word and position in the paragraph or line; correlation between the glyph’s neighbours and its position in the paragraph or line; and certain differences between the scribes. It concludes that no single, simple mechanic can be responsible for all these, and that further study should aim to explain them with multiple mechanisms.

Session Four – 1 December 2022, 1530-1600 CET

Farley Katz

From Voynich to the Beinecke, the Trail of Ownership.

Abstract:

The ownership history of the “Voynich Manuscript” between Wilfred Voynich’s death in 1930 and its donation to the Yale Beinecke Library in 1969 is not well understood because key documents were not available or were misunderstood. Voynich’s will and parts of his wife Ethel’s probate file have now been located and posted online. These documents, together with Ethel’s will, allow its ownership to be traced. In his will, Voynich left a 60 percent interest in the manuscript to Ethel and 40 percent to Anne Nill, his office manager and Ethel’s life-long companion. Voynich expressed his desire that the manuscript be sold to an institution, not an individual, and requested Ethel, Anne and three professors to form a committee to help achieve that goal. After his death, Anne continued to operate Voynich’s rare book business, but business was bleak in the Great Depression and the manuscript did not sell. In 1960, Ethel died and left her interest in the manuscript to Winifred Gaye, a dear friend whom she viewed as her daughter. In 1961, before the estate was distributed, Anne, acting as executrix of the estate and on her own behalf, sold the manuscript to H.P. Kraus for \$24,500. Kraus, one of the most important booksellers of the twentieth century, tried to sell the manuscript for as much as \$160,000. In 1969 he abruptly donated it to the Beinecke Library, apparently to obtain significant tax benefits that would be eliminated after that year.

Session Four – 1 December 2022, 1600-1630 CET

Klaus Schmeh and Elonka Dunin

The Voynich Manuscript Compared with Other Encrypted Books.

Abstract:

The Voynich Manuscript is definitely not the only encrypted book known to crypto historians. The authors of this submission are aware of over 100 other tomes that are written completely or to a considerable part in cipher. This paper provides an overview on these with a focus on their properties and the reasons why they were created. The main goal of this paper is to compare the Voynich Manuscript with other encrypted books in order to draw conclusions and make educated guesses with respect to some of the unanswered questions surrounding this famous cryptogram. Most of all, after we categorize the other known encrypted books, then based on the Voynich Manuscript's similarities with books in various categories, the purpose of the manuscript can be narrowed down with some certainty to two categories: the knowledge book category and the hoax category.

Session Four – 1 December 2022, 1630-1700 CET

Stefan Guzy

Book transactions of Emperor Rudolf II 1576-1612. New findings on the earliest ownership of the Voynich manuscript.

Abstract:

Emperor Rudolf II, who in a 17th century letter is stated to have bought the codex for 600 ducats, is regarded as the first owner of the Voynich manuscript, though further evidence is missing. Due to the large amount of money involved, it is safe to assume that some written notes must have been taken for accounting purposes by the imperial authorities. This study therefore considers the different ways this transaction could have been executed by the imperial administration. It also explores where such an expensive book could have been stored – most probably in the Kunstkammer, the Emperor's personal collection of rare and precious things. In the period of Rudolf's reign, from 1576 to 1612, a total of 126 transactions of books could be found from analysing approximately 6900 journal entries. A transaction in 1599, involving a small number of alchemical manuscripts bought for 600 florin from the Augsburg physician Carl Widemann, seems to present a way in which the Voynich manuscript could have got into the Emperor's collection. Additionally, Widemann's friendship with the botanist and Orient traveller Leonhard Rauwolf sheds possible light on how he could have acquired the manuscript, if he was indeed the owner.

Keynote – 1 December 2022, 1700-1750 CET

Lisa Fagin Davis

Voynich Paleography.

Abstract:

By applying the principles of Latin Paleography to the Voynichese writing system, it is possible to identify distinguishing features of individual scribes. This lecture will explain the principles of Latin paleography and demonstrate how these principles can be profitably applied to the writing system of the Voynich manuscript in order to 1) distinguish different scribal hands; 2) identify the scribal output of each hand; and 3) potentially identify abbreviations and ligatures in the writing system. Finally, the speaker will explain the implications of these conclusions for current and future research on the Voynich Manuscript.