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

Daniel E. Gaskell [1] and Claire L. Bowern [1]

[1] *Yale University, PO Box 208236, New Haven, CT 06520, USA*

**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-long-long 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.

**Keywords**

Voynich manuscript, gibberish, hoax hypothesis

## 1. Introduction

The Voynich Manuscript (VMS, Beinecke MS 408) is an undeciphered codex believed to have been created in the 15th century [1]. Since coming to worldwide attention in the early 1900s, a central point of debate has been whether the VMS encodes meaningful information or should instead be regarded as a medieval hoax, essentially devoid of linguistic meaning [2]. Central to this debate are the peculiar statistical properties of the VMS text (or "Voynichese"). On the one hand, the VMS text exhibits a number of properties which are not typically observed in natural language, such as low conditional character entropy [3] and a high degree of similarity between adjacent words [4]. On the other hand, the VMS text exhibits other properties thought to be indicative of natural language, such as obeying Zipf's law [5] and showing regularities in word morphology [6]. Network analyses have also identified patterns of "topic words" and large-scale information structures which resemble meaningful documents [5–7]. Proponents of the "meaningful text" hypothesis therefore argue that the VMS text is too

International Conference on the Voynich Manuscript 2022, November 30--December 1, 2022, University of Malta.
EMAIL: daniel.gaskell@yale.edu (A. 1); claire.bowern@yale.edu (A. 2)
ORCID: 0000-0002-0306-7943 (A. 1); 0000-0002-9512-4393 (A. 2)

CEUR Workshop Proceedings (CEUR-WS.org)

nonrandom to be meaningless [8], while proponents of the "hoax" (or "gibberish") hypothesis counter that the VMS text does not resemble any known language or cipher system [9].

The implicit assumption underlying much of this controversy is that if the VMS were genuinely meaningless, it would appear random. Several authors have challenged this assumption by proposing generative algorithms which could be used to produce large quantities of text that is meaningless but still significantly nonrandom [10, 11]. Here we challenge this assumption further. If the VMS is indeed a hoax, it is likely that the VMS scribe(s) would have generated the text intuitively rather than by relying on an explicit generative algorithm. That is, rather than rolling dice to generate pure random noise or following an explicit set of rules, the scribe(s) simply invented words and phrases as they wrote in order to create the intuitive appearance of language. Because literate humans have an intuitive sense of what written language looks like, the resulting text could, in theory, be significantly non-random. In order to determine whether the VMS is too non-random to be meaningless, therefore, we must compare it to real human-produced samples of meaningless text, not simply to random noise.

This study attempts to establish a baseline for the properties of such text by collecting real samples of human-produced gibberish. We address four main questions: 1) How well can intuitive gibberish replicate the properties of meaningful text? 2) How well can intuitive gibberish replicate the properties of the VMS? 3) Are there markers which can be used to distinguish meaningful from meaningless texts? 4) In aggregate, does the VMS more closely resemble meaningful or meaningless text?

## 2. Methods

We recruited 42 volunteers to generate short (1–3 page), handwritten documents containing meaningless text, and compared them to the VMS text and a corpus of documents known to be meaningful. Thirty-nine participants were students participating in a Yale undergraduate course on the VMS taught by author C.L.B. (19 from the 2018 class, where the exercise was carried out before students had become familiar with the statistical properties of the VMS text, and 20 from the 2019 class, where the exercise was carried out later in the semester). An additional three participants were recruited from the public and had no knowledge of the study's connection to the VMS. Three writing samples were also contributed by the authors. The impact of excluding one or more of these groups is considered below. All participants spoke fluent English, but a range of other language backgrounds are also represented, including Latin, Hebrew, Spanish, Cantonese, and Australian Indigenous languages.

Protocol documents for reproducing the exercise and a complete archive of results are available on the project's GitHub repository at https://github.com/danielgaskell/voynich. Volunteers were given an instruction sheet telling them to "create a 'document' by filling three pages with fake, meaningless text in a 'language' that you make up as you go. Ideally, this 'language' should not actually mean anything, but should appear realistic enough that most observers would not be able to distinguish it from a real language they simply did not know." In order to better replicate the scribal conditions of the VMS, participants were instructed to write in pen without punctuation, and to use the lowercase Latin alphabet to facilitate transcription. Volunteers were also asked whether they "would consider [themselves] to have specialist knowledge in linguistics and/or conlanging." Completed documents were transcribed into Unicode text documents, preserving line breaks, and analyzed as described below.[1]

For comparison with our gibberish samples, we compiled a corpus of 75 meaningful texts (5.6 million words) including both ancient and modern languages across multiple families, literary and technical documents, and natural and constructed languages (full list available on GitHub). Documents were stored in Unicode text format, preserving original line breaks where possible or otherwise word-wrapping to a column width of 60 characters. Because the appropriate transcription and interpretation of the VMS glyphs is uncertain [1, 3], we included five different transcriptions or subsets in our VMS corpus: Glen Claston's minimally-decomposed v101 transcription [12]; Takeshi Takahashi's maximally-decomposed EVA Full and EVA Basic transcriptions [13]; and the Currier A and B subsets

[1] It is important to note that, because the generative approach which may have been followed by the VMS scribe(s) is unknown, we intentionally refrained from giving participants an explicit algorithm to follow when generating their gibberish. Our intent was to capture the distribution of outcomes which *can* occur in gibberish documents, rather than simulating a specific algorithm which may or may not have been employed. These data therefore allow us to test whether it is possible for the VMS to be gibberish, but not necessarily to identify the exact method used. It is also important to note that participants did not simply develop their own encipherment systems for meaningful text, but were instructed to generate documents with no linguistic meaning at all—albeit ones intended to have the appearance of linguistic meaning.

in EVA Basic, as identified in Jorge Stolfi's interlinear file v16e6 [14]. This allows us to treat the properties of the VMS as a distribution of possibilities rather than assuming which method is correct.

A total of 42 statistical parameters were calculated for each document, summarized in Table 1. To neutralize the effects of sample length, variables were calculated on randomized 200-word excerpts from each document, taking the mean value over 100 iterations. Variables yielding two-tailed distributions (e.g., word lengths) were described using three parameters: mean, standard deviation, and skew. Variables yielding one-tailed distributions (e.g., rank-ordered character frequencies) were described using two parameters: the maximum observed value and a shape parameter $\beta$ obtained by sorting values in descending rank order and fitting the function $f(x) = e^{\frac{-x}{\beta}}\beta^{-1}$, where $x$ is rank.

We assessed whether variable means were different between groups using Welch's bootstrapped t-tests where $\alpha = 0.05$. We calculated a rough metric for the probability of documents in one class resembling documents in another class by using kernel density estimation (KDE) to obtain a probability density function for the first class of documents and assessing what proportion of the area fell within the 5–95% quantiles of the second class of documents. KDE was performed using the "density" function in R version 4.1.0 [15] with bandwidth estimated after Silverman's rule of thumb [16]. We assess that it is *plausible* for the first document class to resemble the second if this overlap is at least 5%; we assess that it is *probable* for the first document class to resemble the second if this overlap is at least 50%.

To assess whether the VMS text more closely resembles the gibberish or meaningful texts in our dataset, we employed random forest classification using the R package randomForest version 4.6-14 [17]. Random forest classification is a common machine-learning approach to automated classification that is robust to overfitting and the presence of large numbers of colinear variables [18].

**Table 1**
Description of statistical metrics

| Variable | Definition |
|---|---|
| charbias_mean<br>charbias_std<br>charbias_skew | Distribution of coefficients of variance on a 10-bin heatmap of how often each character appears at a given position in a line, duplicated by the number of times that character is used to avoid giving too much weight to rare characters. Higher charbias_mean = characters are more biased towards particular positions in a line. |
| charbias_words_mean<br>charbias_words_std<br>charbias_words_skew | Equivalent to charbias_mean etc. above, but using a 5-bin heatmap of positions within words. Higher charbias_words_mean = characters are more biased towards particular positions in a word. |
| chardist_max<br>chardist_shape | Distribution of character frequencies. Higher chardist_shape = smaller difference between the frequencies of more- and less-used characters. |
| compression | % of original length after compressing the entire sample with the DEFLATE algorithm. Higher = greater total information content. |
| entropy | 2nd-order conditional character entropy. Higher = less predictable text. |
| flipped_pairs | Proportion of 2-word pairs which also appear in reversed order. |
| ngramdist_max<br>ngramdist_shape | Distribution of frequencies of 1-, 2-, and 3-character sequences. Higher ngramdist_shape = smaller difference between the frequencies of more- and less-used character sequences. |
| repeated_chars | Proportion of repeated characters in a sample (e.g., *aa*). |
| tripled_chars | Proportion of tripled characters in a sample (e.g., *aaa*). |
| repeated_words | Proportion of repeated words in a sample (e.g., *qokeey qokeey*). |
| tripled_words | Proportion of tripled words in a sample (e.g., *qokeey qokeey qokeey*). |
| unique_chars | Number of unique characters used in a sample. |

| | |
|---|---|
| unique_ngrams | Number of unique 1-, 2-, and 3-character sequences used in a sample. |
| unique_words | Number of unique words used in a sample, after cleaning. |
| wordbias_lines_mean<br>wordbias_lines_std<br>wordbias_lines_skew | Equivalent to charbias_mean etc. above, but using a 5-bin heatmap and counting words instead of characters. Higher wordbias_lines_mean = words are more biased towards particular positions in a line. |
| wordbias_mean<br>wordbias_std<br>wordbias_skew | Equivalent to charbias_mean etc. above, but using a 5-bin heatmap of the entire sample (bin 1 = start of sample, bin 5 = end of sample) and counting words instead of characters. Higher wordbias_mean = words are more biased towards particular positions in a 200-word sample. |
| wordchange_mean<br>wordchange_std<br>wordchange_skew | Distribution of the Levenshtein distance between each word and the word which appears immediately before it, divided by its length and wordunique_mean. Higher wordchange_mean = adjacent words are more different from one another relative to the average. |
| worddist_max<br>worddist_shape | Distribution of word frequencies. Higher worddist_shape = smaller difference between the frequencies of more- and less-used words. |
| wordlen_autocorr | Moran's autocorrelation statistic *I* applied to a 1-D vector of word lengths. Higher = the lengths of adjacent words tend to be more similar. |
| wordlen_mean<br>wordlen_std<br>wordlen_skew | Distribution of the lengths of words in the sample. Higher wordlen_mean = longer average length of words. |
| wordlen_unique_mean<br>wordlen_unique_std<br>wordlen_unique_skew | Distribution of the lengths of unique words in the sample. Higher wordlen_unique_mean = longer average length of unique words. |
| wordunique_mean<br>wordunique_std<br>wordunique_skew | Distribution of the mean Levenshtein distance between vocabulary word and every other vocabulary word, divided by word length. Higher wordunique_mean = vocabulary is more structurally diverse. |
| zipf | LMZ statistic for how closely word frequencies obey Zipf's law [19]. |

## 3. Results

Results are summarized in Figure 1. We find that gibberish documents vary widely in their statistical properties, but are capable of replicating the properties of meaningful text across every variable tested (36 probable, 6 plausible). However, mean values do differ between gibberish and meaningful texts to a statistically significant degree ($p < 0.05$). Compared to meaningful texts, gibberish has lower mean information content (compression); lower mean conditional character entropy (entropy); higher mean occurrences of repeated characters and words (repeated_chars, repeated_words); higher mean bias in where characters appear in a line (charbias_mean) and where words appear in a 200-word section (wordbias_mean); higher mean autocorrelation of word lengths (wordlen_autocorr; see below); and obeys Zipf's law less precisely (zipf). Gibberish documents were also more likely to exhibit high numbers of tripled characters and words (tripled_chars, tripled_words), although mean occurrence rates were not significantly different. All of these tendencies are also significantly observed in the VMS, with the exception of charbias_mean, which is not significantly different between the VMS and meaningful text. The VMS plausibly resembles meaningful text for 38 out of 42 metrics, while it plausibly resembles gibberish for 41 out of 42, with the exception being charbias_words_mean (see Discussion). No significant differences in means were observed between specialist and nonspecialist participants.

Random forest classification identified all five samples in our VMS corpus as more closely resembling gibberish than meaningful text, albeit with low confidence (Figure 2). The ten variables identified as most important for distinguishing groups were, from most to least important, charbias_mean, wordbias_mean, wordbias_skew, wordbias_std, compression, repeated_words, repeated_chars, charbias_words_skew, charbias_std, and wordlen_autocorr.

A notable feature of the VMS that has to our knowledge only been discussed by one other publication [20] is positive autocorrelation of word lengths. Word lengths in most meaningful texts are negatively autocorrelated: that is, long words tend to be interspersed with short words (long-short-long-short). By contrast, the VMS exhibits positive autocorrelation (long-long-short-short). Positive autocorrelation is only observed in a limited number of natural languages, but is common in gibberish (Figure 3).

Because a subset of participants (the 2019 class and the authors) possessed some prior knowledge of the properties of the VMS, there is a possibility of these participants attempting (consciously or unconsciously) to mimic the VMS. We found that omitting the authors' samples did not alter any of the observations described above. When comparing the 2019 class against the 2018 class (which differed in their degree of background knowledge), the only statistically significant differences were 11% higher charbias_mean, 26% lower charbias_skew, 10% lower chardist_shape, and 6% higher unique_chars. As these changes are either small or involve variables unimportant to the foregoing analysis, we conclude that data contamination is unlikely to substantially invalidate our results.

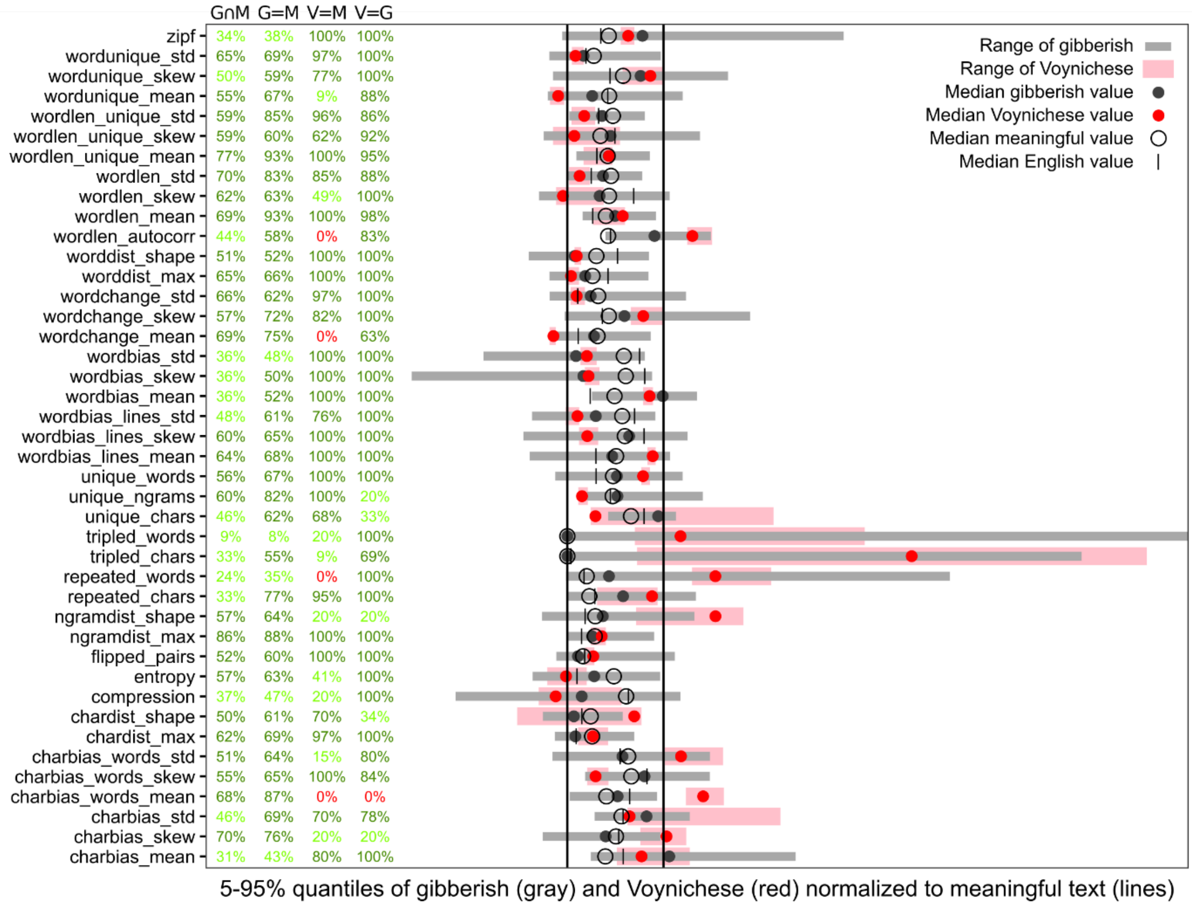


**Figure 1**: Summary of results. **Rows**: For each parameter, values are scaled such that the tall vertical lines indicate the 5% and 95% quantiles of the distributions observed in meaningful texts. 5–95% quantiles of gibberish (gray) and Voynichese (red) are shown as horizontal bars. The placement of each bar relative to the lines thus indicates how the distribution of values differs from the distribution observed in meaningful texts. (Note that tripled_words continues beyond the right plot margin by 2.3x the plot width.) **Columns**: From left to right: 1) Jaccard index of similarity between density areas of gibberish and meaningful texts; 2) probability that a gibberish sample falls within the 5–95% quantiles of meaningful text; 3) probability that a VMS transcription falls within the 5–95% quantiles of meaningful text; 4) probability that a VMS transcription falls within the 5–95% quantiles of gibberish.

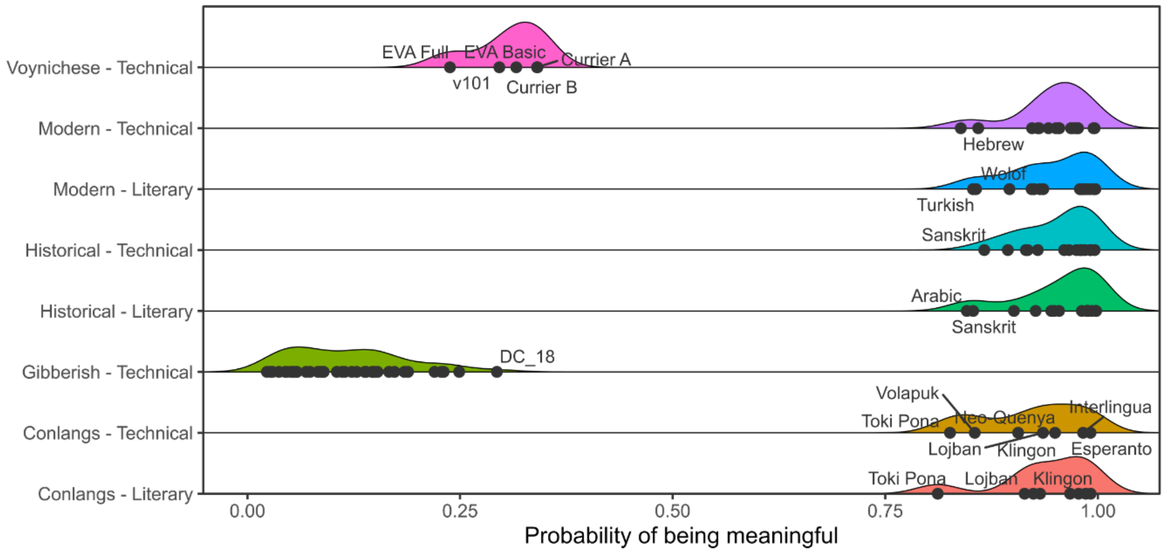


**Figure 2**: Random forest classification of corpus. Points indicate individual documents, with kernel density estimates as colored regions. VMS samples are all <0.5, indicating that they more closely resemble gibberish than meaningful text. (The gibberish sample which the model classified as most resembling meaningful text, DC_18, was produced by a nonspecialist member of the 2018 class.)

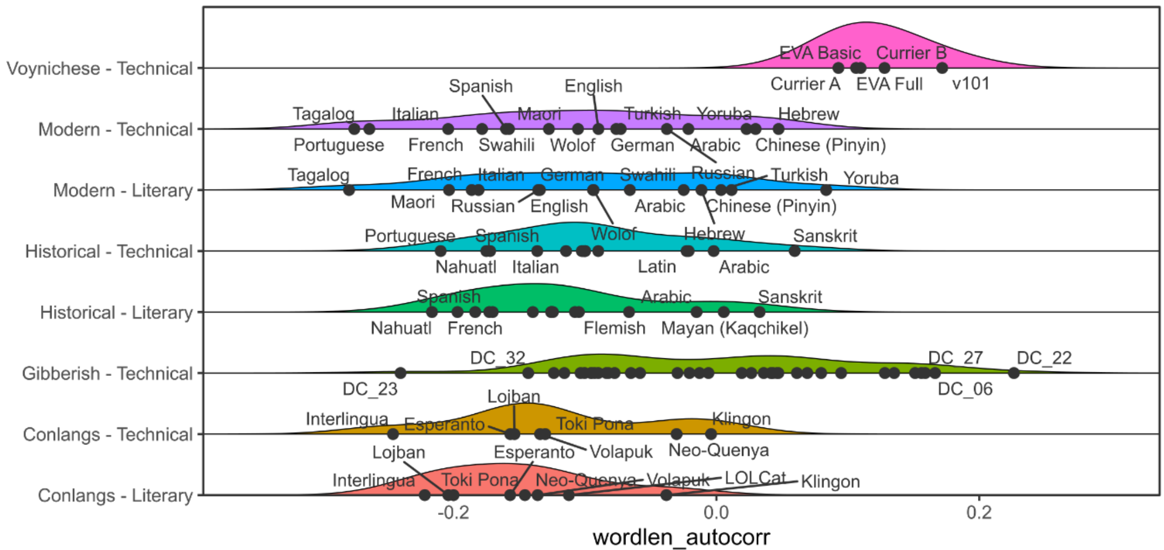


**Figure 3**: Autocorrelation of word lengths by document category.

## 3.1. Visual properties

Samples varied greatly in layout, from uninterrupted streams of text to carefully formatted pseudo-documents with illustrations, labeled diagrams, and text arranged in stanzas or geometric patterns. Unexpectedly for meaningless texts, some samples contained strikeouts or corrections. The VMS notably contains few corrections, which has been argued to indicate either that it was a copy [21] or that the scribe(s) did not care about the meaning [22].[2] A majority of participants maintained an English-like sense of vowel usage, yielding text that could subjectively be considered pronounceable.

We observe that in some portions of the VMS, text is wrapped around the illustrations in a way which appears to have required the scribe(s) to have chosen words of specific lengths to fit (e.g., the

[2] A further experiment in which the same participants copied text in an unknown language found that the urge to strikeout and correct mistakes was insuppressible, even when participants were requested not to do so. It is unknown whether this can be generalized to 15th-century copyists.

words between the stems on VMS 29v, Figure 4). It is possible that the VMS scribe(s) simply employed nonstandard word-breaks in these cases; alternatively, we hypothesized that in the absence of linguistic meaning which would pre-determine word lengths, gibberish writers might be inclined to self-select word lengths which improved the text wrapping. To test this hypothesis, participants from the 2018 class were encouraged to write their samples on stationery containing a selection of plant illustrations, created by masking out the text from scans of the herbal section in the VMS. Other participants were allowed to draw their own illustrations if desired. In total, 31 samples included illustrations, of which we assessed that 17 showed some evidence of word lengths being selected to improve text wrapping (e.g., Figure 4; assessed by asking, "if the order of words on the page were randomized, would the precision of text wrapping noticeably deteriorate?"). While this finding is difficult to quantify rigorously, these observations suggest that the text wrapping features observed in the VMS may also be common in gibberish documents. However, we are not aware of any work investigating the properties of the Voynichese words which are wrapped around illustrations.

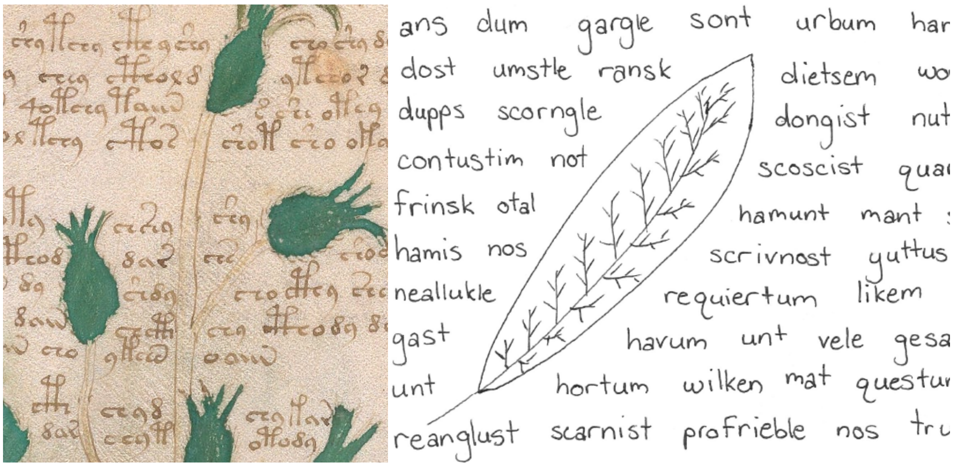


**Figure 4**: Comparison of Voynichese (left, VMS 29v) with gibberish from one of our participants (right), showing potentially analogous selection of word lengths to wrap the text around the illustration.

## 4. Discussion

We cannot and do not attempt here to prove that the VMS is gibberish. Our results do, however, invalidate traditional arguments that the small-scale structure of Voynichese is too non-random to be meaningless [1, 5, 23–25]. Even without an explicit generative algorithm, intuitively-generated gibberish can both adequately replicate many features of natural language and produce a wide range of other unusual properties, depending on the author and text in question. If the VMS is indeed gibberish, in other words, it is perhaps not surprising that it would be both language-like and statistically unusual.

In statistics, the concept of a "null model" refers to the simplest explanation for a dataset against which other models are tested. Much VMS research to date has implicitly used the null model of pure randomness, effectively arguing that, because the VMS is less random than rolling dice, it must be meaningful. We propose that a more appropriate null model is that the VMS was generated by intuitive gibberish such as that produced here. In order to prove that the VMS is meaningful, in other words, we must show that it is less random than human-produced gibberish, not simply that it is less random than rolling dice. While our results cannot conclusively establish whether the VMS is gibberish or not, they help to establish the baseline of what such gibberish might actually look like.

The sole VMS metric which our gibberish samples are unable to replicate is the VMS's unusually large bias in character placement within words (charbias_words_mean). This is likely related to a well-documented feature of the VMS in which certain glyphs appear almost exclusively at the start or end of words [26]. If the VMS is meaningful, these may represent suffix abbreviations, positional variation in encipherment, or typographical flourishes; if the VMS is meaningless, the scribe(s) may have used these glyphs in these positions simply because they were visually appealing. (For example, consider the difference between and .) In either case, the explanation may involve typographic considerations which cannot be tested rigorously using texts restricted to the lowercase Latin alphabet.

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 [5–7]. At present, these features pose a serious challenge to proponents of the hoax hypothesis. However, while it is premature to assume that gibberish can replicate these features, it is equally premature to assume that it cannot; in theory, the properties of a scribe's gibberish might drift considerably over the course of the weeks or months required to generate a VMS-length manuscript, introducing significant large-scale nonrandomness. If the scribe took breaks between sections, or only kept out material from the current section to reference when copying vocabulary, further spatial patterns might arise.[3] Insofar as possible, our results appear consistent with this hypothesis. The presence of "topic words" may be consistent with some participants' reports that they made conscious decisions about word associations, even if the exact meanings of those words remained undefined; for instance, one 2018 participant captioned an illustration "pitshol" and then used the words "pitshol" or "pitsholh" repeatedly in the paragraphs surrounding it. Word distribution biases in our samples also resemble or exceed meaningful texts at the 200-word scale (wordbias_mean, Figure 1). However, whether such simple processes can reproduce the structure observed across pages and sections of the VMS [5–7] remains to be tested. It is also unknown whether such processes might tend to introduce too much spatial variation, destroying any recognizable consistency of the gibberish.

A viable approach to future work (albeit one which introduces its own uncertainties) may be to use the lessons learned here to construct automated algorithms which can generate larger volumes of gibberish, similar to the method of Timm and Schinner [11]. Our results are generally consistent with the proposal of Timm and Schinner that the VMS was generated by a process of "self-citation": that is, that the VMS scribe(s) generated the text largely by copying or modifying words appearing earlier in the same section [11]. This process is argued by these authors to explain both the finer- and larger-scale features of the VMS [11]. Informal interviews and class discussions confirmed that many participants did indeed adopt this type of approach 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. We hypothesize that such processes of self-citation and self-correction may explain many of the unusual features of gibberish. For instance, some participants reported a tendency to write a series of long words, realize they had not written any short words recently, and then self-correct by switching to short words. This tendency may be responsible for the observed positive autocorrelation of word lengths. Elsewhere, inaccurate perspectives on the prevalence of certain language features (such as word repetition or the reuse of prefixes and suffixes) may have led to participants generating vocabulary that felt realistic to them, but was more homogeneous and lower in information content than real language. By incorporating these observations of human behavior into automated text-generating algorithms, future work may be able to more effectively test whether the VMS is meaningful.

A remaining possibility is that the VMS encodes meaningful information, but it is concealed by steganography within a larger body of gibberish. The use of gibberish to generate cryptographic nulls could explain why the VMS exhibits gibberish-like statistical properties without necessarily requiring the text to be meaningless. However, we leave this intriguing possibility for future researchers.

## 5. Acknowledgements

We thank the study participants, as well as four anonymous reviewers for their helpful comments.

## 6. References

[1] M. D'Imperio, *The Voynich Manuscript: An Elegant Enigma*. Fort George G. Meade, Md.: National Security Agency/Central Security Service, 1978.

[2] R. Clemens, Ed., *The Voynich Manuscript:* Illustrated Edition. New Haven; London: Yale University Press, 2016.

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

[3] To take this idea to its extreme, it is interesting to consider whether the different hands in the VMS attributed to multiple scribes [26, 27] could plausibly result from a single scribe taking a long enough break to partially forget their own script and vocabulary.

[4] M. Barlow, "The Voynich Manuscript - by Voynich?," *Cryptologia*, vol. 10, no. 4, pp. 210–216, Oct. 1986, doi: 10.1080/0161-118691861010.

[5] M. A. Montemurro and D. H. Zanette, "Keywords and Co-Occurrence Patterns in the Voynich Manuscript: An Information-Theoretic Analysis," *PLOS ONE*, vol. 8, no. 6, p. e66344, Jun. 2013, doi: 10.1371/journal.pone.0066344.

[6] S. Reddy and K. Knight, "What We Know About The Voynich Manuscript," in *Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, Portland, OR, USA, Jun. 2011, pp. 78–86.

[7] R. Sterneck, A. Polish, and C. Bowern, "Topic Modeling in the Voynich Manuscript." arXiv, Jul. 06, 2021. doi: 10.48550/arXiv.2107.02858.

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

[9] T. Timm and A. Schinner, "Review of The linguistics of the Voynich manuscript by Claire Bowern and Luke Lindemann," *Cryptologia*, vol. 45, no. 5, pp. 434–438, Sep. 2021, doi: 10.1080/01611194.2021.1911875.

[10] G. Rugg, "An Elegant Hoax? A Possible Solution to the Voynich Manuscript," *Cryptologia*, vol. 28, no. 1, pp. 31–46, Jan. 2004, doi: 10.1080/0161-110491892755.

[11] T. Timm and A. Schinner, "A possible generating algorithm of the Voynich manuscript," *Cryptologia*, vol. 44, no. 1, pp. 1–19, Jan. 2020, doi: 10.1080/01611194.2019.1596999.

[12] R. Zandbergen, "Voynich MS - Text Analysis - Transliteration of the Text," *Voynich.nu*, 2022. http://www.voynich.nu/transcr.html (accessed Aug. 07, 2022).

[13] T. Takahashi, "Voynich Manuscript - transcription," 2000. http://www.voynich.com/pages/index.htm (accessed Oct. 12, 2016).

[14] J. Stolfi, "Reeds/Landini's interlinear file converted to EVA," *Jorge Stolfi homepage*, Dec. 29, 1998. https://www.ic.unicamp.br/~stolfi/voynich/98-12-28-interln16e6/ (accessed Jul. 28, 2022).

[15] R Core Team, "R: A language and environment for statistical computing." R Foundation for Statistical Computing, Vienna, Austria, 2016. [Online]. Available: https://www.R-project.org/

[16] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. Boca Raton: Chapman and Hall, 1986.

[17] A. Liaw and M. Wiener, "Classification and Regression by randomForest," *R News*, vol. 2, no. 3, pp. 18–22, 2002.

[18] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Second Edition. Springer, 2009.

[19] C. M. Urzúa, "A simple and efficient test for Zipf's law," *Economics Letters*, vol. 66, no. 3, pp. 257–260, Mar. 2000, doi: 10.1016/S0165-1765(99)00215-3.

[20] V. Matlach, B. A. Janečková, and D. Dostál, "The Voynich manuscript: Symbol roles revisited," *PLOS ONE*, vol. 17, no. 1, p. e0260948, Jan. 2022, doi: 10.1371/journal.pone.0260948.

[21] J. H. Tiltman, "The Voynich Manuscript: 'The Most Mysterious Manuscript in the World.'" National Security Administration, DOCID: 631091, 1968.

[22] G. Rugg and G. Taylor, "Hoaxing statistical features of the Voynich Manuscript," *Cryptologia*, vol. 41, no. 3, pp. 247–268, May 2017, doi: 10.1080/01611194.2016.1206753.

[23] G. Landini, "Evidence of Linguistic Structure in the Voynich Manuscript Using Spectral Analysis," *Cryptologia*, vol. 25, no. 4, pp. 275–295, Oct. 2001, doi: 10.1080/0161-110191889932.

[24] D. R. Amancio, E. G. Altmann, D. Rybski, O. N. O. Jr, and L. da F. Costa, "Probing the Statistical Properties of Unknown Texts: Application to the Voynich Manuscript," *PLOS ONE*, vol. 8, no. 7, p. e67310, Jul. 2013, doi: 10.1371/journal.pone.0067310.

[25] K. Schmeh, "A Milestone in Voynich Manuscript Research: Voynich 100 Conference in Monte Porzio Catone, Italy," *Cryptologia*, vol. 37, no. 3, pp. 193–203, Jul. 2013, doi: 10.1080/01611194.2013.797045.

[26] P. Currier, "Some Important New Statistical Findings," in *New Research on the Voynich Manuscript: Proceedings of a Seminar*, Washington, D.C., Nov. 1976.

[27] L. F. Davis, "How Many Glyphs and How Many Scribes? Digital Paleography and the Voynich Manuscript," *Manuscript Studies: A Journal of the Schoenberg Institute for Manuscript Studies*, vol. 5, no. 1, pp. 164–180, 2020, doi: 10.1353/mns.2020.0011.