

---

# The strange resonances of the Voynich manuscript

**Stephan Vonfelt**

*stephan.vonfelt@free.fr*

---

*ABSTRACT. The Voynich manuscript, written in an unknown language, is generally attributed to medieval Europe, through its illustrations tinged with realism and fantastic. In the light of enigmatic characters, the composition is normal, however the chronology presents unknown correlations among European languages, partly found in Chinese. The long memory phenomenon could arise from a fractional stochastic process. But these resonances, also encountered on semantic and binary levels, make the Voynich manuscript a complex construction, manifestly out of reach of a medieval scholar, a fortiori of a forger pressed by money. The most likely hypothesis remains the one of an alchemist, relying on a naturally rhythmic language to pass on an incantation or an initiation. Nevertheless, the question remains open...*

**KEYWORDS :** *Voynich manuscript, cryptography, stylometry, rhythm, time series, correlation, long memory, fractional process.*

**RÉSUMÉ.** *Le manuscrit de Voynich, écrit dans une langue inconnue, est généralement attribué à l'Europe médiévale, à travers ses illustrations teintées de réalisme et de fantastique. À l'aune de caractères énigmatiques, la composition est normale, cependant la chronologie présente des corrélations inédites parmi les langues européennes, partiellement retrouvées dans le chinois. Le phénomène à mémoire longue pourrait naître d'un processus stochastique fractionnaire. Mais ces résonances, rencontrées par ailleurs sur les plans sémantiques et binaires, font du manuscrit de Voynich une construction complexe, manifestement hors de la portée d'un savant du Moyen Âge, a fortiori d'un faussaire pressé par l'argent. L'hypothèse la plus vraisemblable reste celle d'un alchimiste, s'appuyant sur un langage naturellement rythmé pour transmettre une incantation ou une initiation. Néanmoins, la question reste ouverte...*

**MOTS-CLÉS :** *manuscrit de Voynich, cryptographie, stylométrie, rythme, séries chronologiques, corrélation, mémoire longue, processus fractionnaire.*

---

## 1. Context

### 1.1 The Voynich manuscript

Currently held by the University of Yale<sup>1</sup>, the manuscript owes its name to Wilfrid Voynich, an American bibliophile of Polish origin who acquired the book in 1912. In an old collection build up by Italian Jesuits, he discovered a singular codex, written in an unknown alphabet. Slipped into the pages, a letter from Jan Marek Marci to Athanasius Kircher, dated 1666, indicated that it had belonged to the Germanic emperor Rudolph II. Keen on esotericism and science, the latter would have bought it in 1586 from John Dee, believing in the acquisition of a work of Roger Bacon, the “doctor mirabilis” of the 13th century.

The parchment is composed of 104 folios, against 116 at the origin. Along the pages, the coloured figures are the only key of a work which seems to oscillate between science, alchemy and magic. The illustrations allow one to draw several parts, related to the topics of botany, astronomy and astrology, biology, cosmology, and pharmacology; the last section, only textual, could gather receipts. The text of the manuscript, probably inserted after the illustrations, is generally written from left to right and from top to bottom, forming several paragraphs. Within a line, the separation of words and characters is sometimes questionable.

### 1.2 Studies on the manuscript

First approaches are based on the form. Having received the manuscript from Marci, Kircher saw similarities with the Illyrian alphabet. At the beginning of the 20th century, Newbold believed to have detected micrographs and anagrams, but the thesis was refuted by Manly. Considering the style of the illustrations and text, experts traditionally located the manuscript in Central Europe, at the turning point of 15<sup>th</sup> and 16<sup>th</sup> century. However, Hodgins and carbon 14 dating the parchment around 1420.

The methodical study of the text supposes its transcription. In the middle of the 20<sup>th</sup> century, Petersen initiated the task manually, carried on by Friedman’s First Study Group in order to provide an electronic version. Perceiving two languages and two hands within the folios, Currier undertook a new transcription in 1976. The complex and rare characters analysed by Zandbergen’s EVA alphabet, Takeshi concluded the exercise in 1998.

The digital age opened. The cryptologist Friedman judged later that the undeciphered manuscript was an artifice, and was joined a few years later by Tiltman. More recently, Stolfi compared “Voynichese” to Asian languages, from the distribution of word lengths. Rugg proposed a construction with a set of syllables and a Cardan grid. Finally, Montemurro and Zanette highlighted the thematic distribution of the words of the manuscript.

The studies frequently privilege the characters and their distribution. Bennett noted the low entropy of a text, whose successive units are correlated. Landini’s spectral analysis nevertheless

---

<sup>1</sup> *Voynich Manuscript*, Beineke Rare Book & Manuscript Library, Yale University.

evoked the relationship of Voynichese with natural languages. Through a finer mesh, Schinner observed that the walk of the bits coding the characters was not free, unlike traditional texts.

## 2. Corpus

### 2.1 Voynich text

Takeshi's exhaustive transcription divides the text into several units: paragraphs, circular or radial lines, titles, labels, single letters<sup>2</sup>... Only the paragraphs are retained, these periods appearing more significant. The punctuation of the transcription between paragraphs, lines and words is kept, like rare characters of the extended alphabet, and unspecified characters. Our attention concentrates however on the basic letters, coded from a to z.

### 2.2 Comparative texts

The repetitive Voynichese is first linked to medieval poetry, in rhythm with alliteration, assonance and rhyme. Through Europe and the centuries, our corpus convenes the *Divina Commedia*<sup>3</sup> for Italian, *Parzival* for German<sup>4</sup>, *Perceval ou Le Conte du Graal* for French<sup>5</sup>, *Alliterative Morte Arthure* for English<sup>6</sup>. The text is further compared to more remote sources: beside the references of Antiquity, the Latin *Aeneid*<sup>7</sup> and the Greek *Iliad*<sup>8</sup>, the Asian track is explored, through the Chinese historical memories of *Shih-Ji*<sup>9</sup>.

The letters are reduced to a canonical form, lower case without diacritic. Greek is thus coded on 24 characters, while Chinese is romanised by the pinyin<sup>10</sup>: the sinographs are too numerous to be confronted with the Latin letters, besides this phonetic transcription founds the Chinese measures of Stolfi. Finally, these texts are shortened to line up with the Voynich manuscript and to build the statistics on equivalent sizes, about 200 000 characters including punctuation and spaces.

## 3. Measures

Along the characters, spaces suggest words, but the separations are sometimes uncertain. Moreover, each combination of letters constitutes a sparse population. Thus, our statistics favour a

---

<sup>2</sup> Stolfi J. (1998), "Reeds/Landini's interlinear file in EVA, version 1.6e6", Instituto De Computação, Universidade Federal Fluminense.

<sup>3</sup> Dante (ca 1310), *Divina Commedia*, Project Gutenberg.

<sup>4</sup> Wolfram von Eschenbach (ca 1220), *Parzival*, Bibliotheca Augustana, Universität Augsburg.

<sup>5</sup> Chrétien de Troyes (1181), *Perceval ou Le Conte du Graal*, Laboratoire de français ancien, Université d'Ottawa.

<sup>6</sup> Unkown author (ca 1400), *Alliterative Morte Arthure*, Medieval Institute, University of Rochester.

<sup>7</sup> Virgil (1<sup>st</sup> century before J.C.), *Aeneis*, Perseus Digital Library, Tufts University.

<sup>8</sup> Homer (8<sup>th</sup> century before J.C. ?), *Ἰλιάς*, Perseus Digital Library, Tufts University.

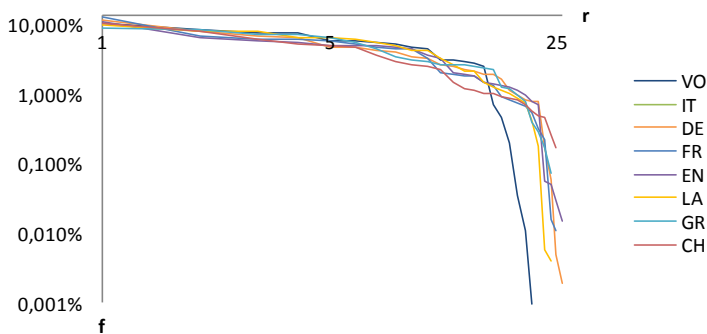
<sup>9</sup> Qian Sima (1<sup>st</sup> century before J.C.), *史記*, Project Gutenberg.

<sup>10</sup> The pinyin tons are ignored, like the above diacritics.

firm and abundant material, the letters<sup>11</sup>. Once these units have been defined, elementary measures analyse the composition of the text, whereas complex measures synthesise its distribution.

### 3.1 Composition

The composition of the text is given by the frequency of each letter. The symbols being heterogeneous (the correspondence, arbitrary for Voynichese, is unsure for Greek), one compares the ranking of letters by decreasing relative frequencies<sup>12</sup>. A logarithmic diagram reveals a pseudo Zipf law, imposed on the whole corpus<sup>13</sup>: on a wide range, the curves merge into one line, indicating a  $f \propto r^{-s}$  type distribution. The Voynich manuscript is characterised only by its concentration, related to rare letters (m, f, g, x, z) or absent letters (b, j, u, v, w).



### 3.2 Distribution

The manuscript having been put into the common law by what precedes, more precise statistics are necessary. Let us note the position of the  $j^{\text{th}}$  occurrence of the letter  $i$  in a text by  $X_i^j$  ( $i \in [1, n_\alpha]$ ,  $j \in [1, n_i]$ ). For a given  $i$ ,  $X_i^j$  is strictly increasing, non-stationary and unsuitable for temporal statistics<sup>14</sup>. The variable is regularised by differentiating:  $Y_i^j = X_i^j - X_i^{j-1}$ . This period measures locally the rarity of a letter, contrary to its density.

#### 3.2.1 Standard deviation

The average  $\mu_i$  of  $Y_i$  is the inverse of the previously analysed frequency. Around this average, the fluctuations are quantified by the standard deviation  $\sigma_i$ . Here again, the comparison of heterogeneous symbols is impossible, which leads to a general standard deviation:  $\sigma = (\sum n_i \sigma_i^2 / \sum n_i)^{1/2}$ ,  $i \in [1, n_\alpha]$ . This new measure does not more dissociate the Voynich manuscript, which is located in the middle of the corpus.



<sup>11</sup> On the effectiveness of the characters compared to the syntactic and semantic levels: Vonfelt (2008), p. 255-256.

<sup>12</sup> The number of occurrences of a letter, divided by the number of characters of the text.

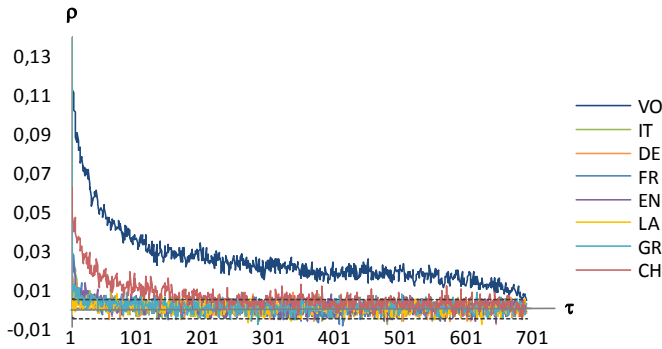
<sup>13</sup> The original Zipf law relates to words, and traces a line of slope -1.

<sup>14</sup> The equivalence of temporal and spatial averages supposes the ergodicity.

### 3.2.2 Time correlations

However, the standard deviation ignores the chronology. The rhythm of the appearances of a letter can be synthesised by the correlation of the variable  $Y_i$  with itself, shifted by  $\tau$  intervals. Ranging between -1 and 1, the coefficient of autocorrelation ensures the comparison of several populations:  $\rho_{Y_i}(\tau) = E(Y_i^j - \mu_i)(Y_i^{j+\tau} - \mu_i) / \sigma_{Y_i}^2$ ,  $j \in [1, n_i]$ . Noticing that this magnitude is equivalent for the reduced variable  $y_i = (Y_i - \mu_i) / \sigma_i$ , and postulating as an initial approach that the reduction places all letters on equal footing, a general correlation  $\rho(\tau)$  is established.

The correlations are fundamentally positive, decreasing and significant down to the confidence interval threshold<sup>15</sup>. The first incursion of Voynichese into this margin is late ( $T \sim 700$ ), closing the window of the observations. On the other hand, the Chinese and a fortiori European invasions are early ( $T_{CH} \sim 80$  and  $T_{EU} \sim 10$ ). Diverse, these curves trace the resonance of a letter in its surround, in other words alliteration and assonance, largo sensu<sup>16</sup>.



The cumulated correlations on the whole field enables to quantify these observations:  $P = \sum \rho(\tau)$ ,  $\tau \in [1, T]$ <sup>17</sup>. Compared to Voynichese, Chinese resounds slightly, whereas European languages produce almost no echo<sup>18</sup>. In this last group, the interpretation of the fluctuations is difficult, but the English prevalence is probably not extraneous to the *Alliterative Morte Arthure*.



<sup>15</sup>  $I = [-2\sigma, 2\sigma]$  with a confidence interval of 95%, where  $\sigma = n^{-0.5}$  from the Barlett formula.

<sup>16</sup> The meanings of these processes vary. Generally, alliteration is the repetition of consonants at the beginning of a word, assonance the repetition of vowels at the end of a verse.

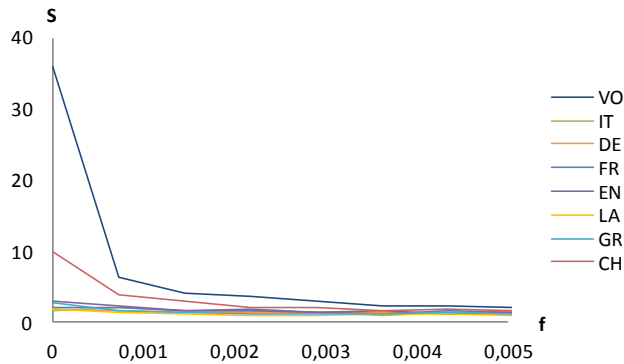
<sup>17</sup> Geometrically, P represents the area under the curve  $\rho(\tau)$ .

<sup>18</sup> Contemporary French does not escape the rule: Vonfelt (2008), p. 234-235.

## 4. Modelling

### 4.1 Memory persistence

Beforehand, it is appropriate to determine whether the memory of the process is short or long. Instead of observing the correlations to infinity, one scans the spectrum at the origin, calculated by the Fourier transform of the function of correlation according to the Wiener-Khintchine theorem. If the value in 0 reflects the preceding resonance<sup>19</sup>, the decrease of the curve forms a pole for Voynichese and Chinese, while European languages remain slack. The process is said to have a long memory in the first case ( $\rho \propto \tau^{-s}$ ), a short memory in the second case ( $\rho \propto e^{-st}$ ).



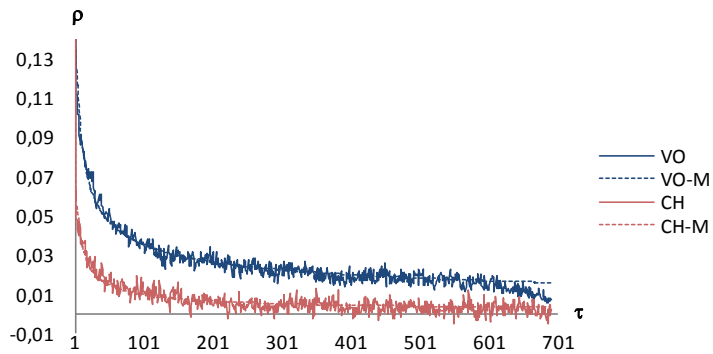
### 4.2 Process

In the first phase, the long term component is eliminated, keeping the series stationary. The Granger and Joyeux method consists in a fractional differentiation of the initial variable to come down to a short memory series:  $Z = (1-L)^d Y$ , where  $L$  is the lag operator; ranging between  $-\frac{1}{2}$  and  $\frac{1}{2}$ ,  $d$  reflects the slope of the previous spectrum on a logarithmic diagram ( $S \propto f^s$ ). In the second phase, the series  $Z$  is treated by the Box and Jenkins “ARMA” process, as a finite linear combination of past values and random events:  $(1-a_1L \dots - a_pL^p) Z = (1-b_1L \dots - b_qL^q)e$ , where  $e$  is a white noise; to keep simplicity,  $p$  and  $q$  will be limited to 1. Finally, the whole constitute an “ARFIMA” process.

European languages are little relevant, and thus excluded at this stage. For Voynichese and Chinese, the long term component is significant ( $d_{VO} = 0.29$ ,  $d_{CH} = 0.20$ ). In the short term, a predictably positive auto-regression ( $\phi_{VO} = 0.27$ ,  $\phi_{CH} = 0.37$ ) is combined with an influent random factor ( $\theta_{VO} = 0.49$ ,  $\theta_{CH} = 0.52$ ). Once these coefficients have been identified, one can confront the theoretical and empirical correlations: on a wide range, the model is faithful. Under distinct features, Voynichese and Chinese seem to find, if not a relative, at least a common gene.

---

<sup>19</sup>  $S(0) = \sum \rho(\tau) = 2P+1$  taking into account the parity of  $\tau$  and its value at 0.



## 5. Conclusion

In the light of its characters, the Voynich manuscript brings out an ordered landscape, alternating dense and sparse surfaces by the effect of correlations. Resonance, significant to a lesser extent for Chinese, is unknown among European languages, which distribute their letters randomly. This long memory phenomenon can be modelled by a fractional random process.

Stolfi's Chinese theory is so revived. Besides, these correlations corroborate the observations of Montemurro and Zanette on the semantic level, like those of Schinner on the binary level. Visibly, the text in question is not the fruit of chance, but of a complex elaboration. With the wealth and the beauty of its illustrations, the manuscript constitutes obviously an accomplished work of art.

These elements undermine the hypothesis of a forger pressed by money. Also, it is difficult to imagine a scholar of the Middle Age handling processes formalised in the 20th century. The author could have relied on a naturally rhythmic language to pass on an incantation or an initiation. The figure of the alchemist addressing a circle of initiates is reborn here, but "the most mysterious manuscript in the world" has probably not said its last word.

## References

- BENETT W. (1976), *Scientific and Engineering Problem Solving with the Computer*, Prentice Hall.
- BOX G. & JENKINS G. (1970), *Time series analysis: Forecasting and control*, Holden-Day.
- CURRIER P. (1976), "Papers on the Voynich Manuscript", in D'Imperio, M.E., *New Research on the Voynich Manuscript: Proceedings of a Seminar*, Washington, D.C.
- D'IMPERIO M. (1978), *The Voynich Manuscript - An elegant enigma*, Aegean Park Press.
- HODGINS G. (2011), *To Crack the Voynich Code*, University of Arizona.
- GRANGER C. & JOYEUX R. (1980), "An introduction to long memory time series and fractional differencing", *Journal of Time Series Analysis*, Wiley, vol. 1.
- JASKIEWICZ G. (2011), "Analysis of Letter Frequency Distribution in the Voynich Manuscript", *Proceedings of the international workshop CS&P*, Warsaw University of Technology.
- LANDINI G. (2001), "Evidence of Linguistic Structure in the Voynich Manuscript using Spectral Analysis", *Cryptologia*, Taylor & Francis, vol. 25, n°4, p. 275-295.
- LANDINI G. & ZANDBERGEN R. (1998), "A Well-kept Secret of Mediaeval Science: the Voynich manuscript", *Aesculapius*, University of Birmingham.
- MANLY J. (1931), "Roger Bacon and the Voynich Manuscript", *Speculum*, Cambridge University Press, vol. 6, p. 345-391.
- MONTEMURRO M. & ZANETTE D. (2013), "Keywords and Co-Occurrence Patterns in the Voynich Manuscript: An Information-Theoretic Analysis", *Plos One*, vol. 8, n°6.
- NEWBOLD W. (1921), "The Cipher of Roger Bacon", *Proceedings of the College of Physicians and Surgeons of Philadelphia*, University of Pennsylvania Press, p. 431-474.
- REEDS J. (1995), "William F. Friedman's Transcription of the Voynich Manuscript", *Cryptologia*, Taylor & Francis, vol. 19, n°1, p. 1-23.
- RUGG G. (2004), "An elegant hoax? A possible solution to the Voynich manuscript", *Cryptologia*, Taylor & Francis, vol. 28, n°1, p. 31-46.
- SCHINNER A. (2007), "The Voynich Manuscript: Evidence of the Hoax hypothesis", *Cryptologia*, Taylor & Francis, vol. 31, n° 2, p. 95-107.
- STOLFI J. (2002), "Chinese Theory Redux: Comparing the VMS and East Asian word length distributions", Instituto De Computação, Universidade Federal Fluminense.
- TILTMAN J. (1967), "The Voynich Manuscript, The Most Mysterious Manuscript in the World", *NSA Technical Journal*, vol. 12, n°3, p.41-85.
- VONFELT S. (2008), *La musique des lettres : variations sur Yourcenar, Tournier et Le Clézio*, Université de Toulouse.
- ZANDBERGEN R. (2012), *The Voynich Manuscript*, [www.voynich.nu](http://www.voynich.nu).