
Les résonances étranges du manuscrit de Voynich

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 unheard correlations among European languages, partly found in Chinese. The long memory phenomenon could arise from a stochastic fractional 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. Contexte

1.1 Le manuscrit de Voynich

Actuellement détenu par l'université de Yale¹, le manuscrit doit son nom à Wilfrid Voynich, un bibliophile américain d'origine polonaise qui acquit le livre en 1912. Dans un fonds ancien constitué par des jésuites italiens, il découvrit un codex singulier, écrit d'un alphabet inconnu. Glissée parmi les pages, une lettre de Jan Marek Marci à Athanasius Kircher, datée de 1666, indiquait qu'il avait appartenu à l'empereur germanique Rodolphe II. Fêré d'ésotérisme et de science, ce dernier l'aurait acheté en 1586 à John Dee, croyant acquérir une œuvre de Roger Bacon, le « doctor mirabilis » du 13^e siècle.

Le parchemin comprend 104 folios, contre 116 à l'origine. Au fil des pages, les figures de couleur sont l'unique clé d'une œuvre qui semble osciller entre science, alchimie et magie. Les illustrations permettent de dégager plusieurs parties, liées aux thèmes de la botanique, l'astronomie et l'astrologie, la biologie, la cosmologie, et la pharmacologie ; la dernière section, uniquement textuelle, pourrait rassembler des recettes. Le texte du manuscrit, vraisemblablement inséré après les illustrations, est généralement orienté de gauche à droite et de haut en bas, formant plusieurs paragraphes. Au sein d'une ligne, la séparation des mots et des caractères est parfois douteuse.

1.2 Études sur le manuscrit

De premières approches se fondent sur la forme. Ayant reçu le manuscrit de Marci, Kircher vit des similarités avec l'alphabet illyrien. Au début du 20^e siècle, Newbold crut déceler des micrographes et des anagrammes, mais la thèse fut réfutée par Manly. A l'examen du style des illustrations et du texte, les experts situèrent traditionnellement le manuscrit en Europe centrale, au tournant du 15^e et 16^e siècle. Cependant, Hodgins et le carbone 14 datent le parchemin vers 1420.

L'étude méthodique du texte suppose sa transcription. Au milieu du 20^e siècle, Petersen initia la tâche manuellement, reprise par le First Study Group de Friedman en vue d'une version électronique. Percevant deux langages et deux mains au sein des folios, Currier entreprit en 1976 une nouvelle transcription. Les caractères complexes et rares analysés par l'alphabet EVA de Zandbergen, Takeshi conclut l'exercice en 1998.

L'ère numérique s'ouvrait. Le cryptologue Friedman jugea que le manuscrit indéchiffré était un artifice, rejoint quelques années plus tard par Tiltman. Plus récemment, Stolfi rapprocha le « voyniche » des langues asiatiques, par la distribution des longueurs des mots. Rugg proposa une construction faite d'un jeu de syllabes et d'une grille de Cardan. Enfin, Montemurro et Zanette mirent en évidence la répartition thématique des mots du manuscrit.

Les études privilégient fréquemment les caractères et leur répartition. Benett nota la faible entropie d'un texte dont les unités successives sont corrélées. L'analyse spectrale de Landini évoqua néanmoins la parenté du voyniche avec les langues naturelles. Par un tamis plus fin,

¹ *Voynich Manuscript*, Beineke Rare Book & Manuscript Library, Yale University.

Schinner observa que la marche des bits codant les caractères n'était pas libre, a contrario des textes traditionnels.

2. Corpus

2.1 Texte de Voynich

La transcription exhaustive de Takeshi découpe le texte en plusieurs unités : paragraphes, lignes circulaires ou radiales, titres, légendes, lettres isolées²... Seuls les paragraphes sont retenus, ces périodes paraissant plus significatives. La ponctuation de la transcription entre les paragraphes, lignes et mots est conservée, ainsi que les caractères rares de l'alphabet étendu, et les caractères indéterminés. Notre regard se concentre cependant sur les lettres de base, codées de a à z.

2.2 Textes comparatifs

Le voyniche répétitif est en premier lieu rapproché de la poésie médiévale, rythmée par l'allitération, l'assonance et la rime. À travers l'Europe et les siècles, notre corpus convoque la *Divina Commedia*³ pour l'italien, *Parzival* pour l'allemand⁴, *Perceval ou Le conte du Graal* pour le français⁵, *Alliterative Morte Arthure* pour l'anglais⁶. Le texte est d'autre part comparé à des sources plus lointaines : hormis les références de l'Antiquité, l'*Énéide* latine⁷ et l'*Illiade* grecque⁸, la piste asiatique est explorée, par les mémoires historiques chinoises du *Shih-Ji*⁹.

Les lettres sont ramenées à une forme canonique, minuscule sans diacritique. Le grec est ainsi codé sur 24 caractères, tandis que le chinois est romanisé par le pinyin¹⁰ : les sinogrammes sont trop nombreux pour être confrontés aux lettres latines, par ailleurs cette transcription phonétique fonde les mesures chinoises de Stolfi. Enfin, ces textes sont abrégés pour s'aligner sur le manuscrit de Voynich et seoir les statistiques sur des effectifs équivalents, de l'ordre de 200 000 caractères avec la ponctuation et les espaces.

3. Mesures

Au fil des caractères, des espaces suggèrent des mots, mais les séparations sont parfois incertaines. Par ailleurs, chaque combinaison de lettres constitue une population clairsemée. Nos statistiques privilégient donc un matériau ferme et abondant, les lettres¹¹. Ces unités définies, des

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

³ Dante (ca 1310), *Divina Commedia*, Project Gutenberg.

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

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

⁶ Auteur inconnu (ca 1400), *Alliterative Morte Arthure*, Medieval Institute, University of Rochester.

⁷ Virgile (1^{er} siècle avant J.C.), *Aeneis*, Perseus Digital Library, Tufts University.

⁸ Homère (8^e siècle avant J.C. ?), *Ἰλιάς*, Perseus Digital Library, Tufts University.

⁹ Qian Sima (1^{er} siècle avant J.C.), *史記*, Project Gutenberg.

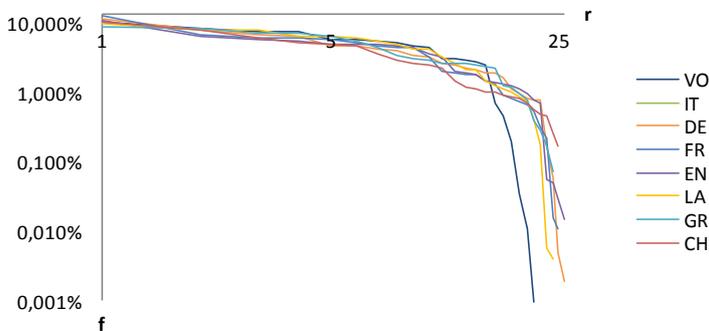
¹⁰ Les tons du pinyin sont ignorés, comme les diacritiques précédents.

¹¹ Sur l'efficacité des caractères par rapport aux niveaux syntaxiques et sémantiques : Vonfelt (2008), p. 255-256.

mesures élémentaires analysent la composition du texte, alors que des mesures complexes synthétisent sa répartition.

3.1 Composition

La composition du texte est donnée par la fréquence de chaque lettre. Les symboles étant hétérogènes (la correspondance est arbitraire pour le voyniche, mal assurée pour le grec), on compare le rangement des lettres par fréquences relatives décroissantes¹². Un diagramme logarithmique révèle une pseudo-loi de Zipf, qui s'impose à l'ensemble du corpus¹³ : sur une large plage, les courbes se confondent en une droite, traduisant une distribution de type $f \propto r^{-s}$. Le manuscrit de Voynich ne se distingue que par sa concentration, liée à des lettres rares (m, f, g, x, z) ou absentes (b, j, u, v, w).

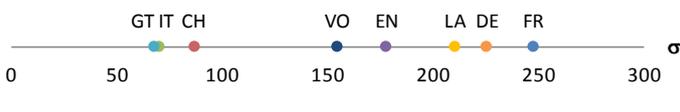


3.2 Répartition

Le manuscrit rangé dans le lot commun par ce qui précède, des statistiques plus précises s'imposent. Notons la position de la j^{e} occurrence de la lettre i dans un texte par X_i^j ($i \in [1, n_\alpha]$, $j \in [1, n_i]$). Pour i donné, X_i^j est strictement croissante, non stationnaire et impropre aux statistiques temporelles¹⁴. La variable est régularisée en différenciant : $Y_i^j = X_i^j - X_i^{j-1}$. Cette période mesure localement la rareté d'une lettre, a contrario de sa densité.

3.2.1 Écart-type

La moyenne μ_i de Y_i est l'inverse de la fréquence déjà analysée. Autour de cette moyenne, les fluctuations sont quantifiées par l'écart-type σ_i . Ici encore, la comparaison de symboles hétérogènes est impossible, d'où le recours à un écart-type général $\sigma = (\sum n_i \sigma_i^2 / \sum n_i)^{1/2}$, $i \in [1, n_\alpha]$. Cette nouvelle grandeur ne démarque pas plus le manuscrit de Voynich, qui campe au milieu du corpus.



¹² Le nombre d'occurrences d'une lettre ramené au nombre de caractères du texte.

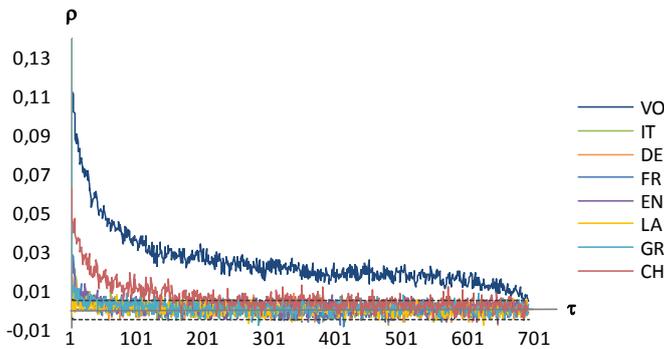
¹³ La loi de Zipf originale porte sur les mots, et trace une droite de pente -1.

¹⁴ L'équivalence des moyennes temporelles et spatiales suppose l'ergodicité.

3.2.2 Corrélation temporelle

L'écart-type ignore cependant la chronologie. Le rythme des apparitions d'une lettre peut se synthétiser par la corrélation de la variable Y_i avec elle-même, décalée de τ intervalles. Compris entre -1 et 1, le coefficient d'auto-corrélation assure la comparaison de plusieurs populations : $\rho_{Y_i}(\tau) = E(Y_i^j - \mu_i)(Y_i^{j+\tau} - \mu_i) / \sigma_{Y_i}^2$, $j \in [1, n_i]$. Remarquant que cette grandeur est équivalente pour la variable réduite $y_i = (Y_i - \mu_i) / \sigma_i$, et postulant en première approche que la réduction place toutes les lettres à égalité, une corrélation générale $\rho(\tau)$ est établie.

Les corrélations sont fondamentalement positives, décroissantes et significatives jusqu'au seuil de l'intervalle de confiance¹⁵. La première incursion du voyniche dans cette marge est tardive ($T \sim 700$), fermant la fenêtre des observations. En revanche, les invasions chinoises et a fortiori européennes sont précoces ($T_{CH} \sim 80$ et $T_{EU} \sim 10$). Diverses, ces courbes tracent la résonance d'une lettre dans son entourage, soit largo sensu l'allitération et l'assonance¹⁶.



Le cumul des corrélations sur l'ensemble du champ permet de quantifier ces observations : $P = \Sigma \rho(\tau)$, $\tau \in [1, T]$ ¹⁷. Par rapport au voyniche, le chinois résonne faiblement, alors que les langues européennes ne produisent quasiment aucun écho¹⁸. Dans ce dernier groupe, l'interprétation des fluctuations est malaisée, mais la prévalence anglaise n'est vraisemblablement pas étrangère à l'*Alliterative Morte Arthure*.



¹⁵ $I = [-2\sigma, 2\sigma]$ avec une confiance de 95%, où $\sigma = n^{-0.5}$ d'après la formule de Bartlett.

¹⁶ Les acceptions de ces procédés varient. Généralement, l'allitération est la répétition de consonnes en début de mot, l'assonance la répétition de voyelles en fin de vers.

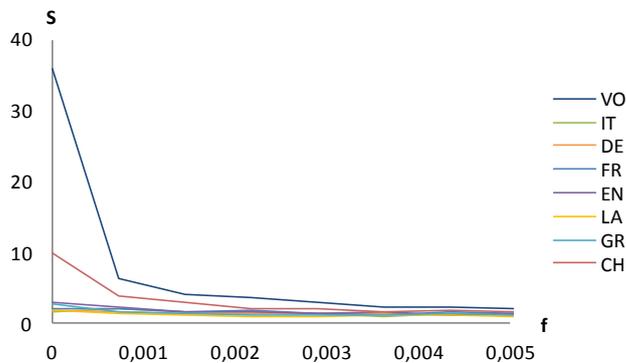
¹⁷ Géométriquement, P représente l'aire sous-tendue par la courbe $\rho(\tau)$.

¹⁸ Le français contemporain n'échappe pas à la règle : Vonfelt (2008), p. 234-235.

4. Modélisation

4.1 Persistance de la mémoire

Préalablement, il convient de déterminer si la mémoire du processus est courte ou longue. Au lieu d'observer les corrélations à l'infini, on scrute le spectre à l'origine, calculé par la transformée de Fourier de la fonction de corrélation selon le théorème de Wiener-Khintchine. Si la valeur en 0 traduit la résonance qui précède¹⁹, la décroissance de la courbe forme un pôle pour le voyniche et le chinois, tandis que les langues européennes restent étales. Le processus est dit à mémoire longue dans le premier cas ($\rho \propto \tau^{-s}$), à mémoire courte dans le second ($\rho \propto e^{-s\tau}$).

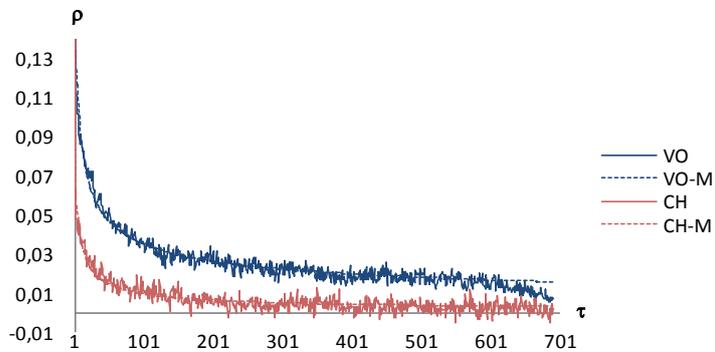


4.2 Processus

Dans une première phase, on élimine la composante à long terme en conservant la stationnarité de la série. La méthode de Granger et Joyeux consiste à différencier la variable initiale par fraction pour se ramener à une série de mémoire courte : $Z = (1-L)^d Y$, où L est l'opérateur de retard ; compris entre $-\frac{1}{2}$ et $\frac{1}{2}$, d reflète la pente du spectre précédent sur un diagramme logarithmique ($S \propto f^s$). Dans une seconde phase, la série Z est traitée par le processus « ARMA » de Box et Jenkins, comme une combinaison linéaire finie de ses valeurs passées et d'événements aléatoires : $(1-a_1L \dots -a_pL^p)Z = (1-b_1L \dots -b_qL^q)e$, où e est un bruit blanc ; pour rester simple, p et q seront limités à 1. Finalement, l'ensemble des phases constitue un processus « ARFIMA ».

Les langues européennes, peu pertinentes, sont exclues à ce stade. Quant au voyniche et au chinois, la composante à long terme est significative ($d_{VO} = 0.29$, $d_{CH} = 0.20$). À court terme, une auto-régression évidemment positive ($\phi_{VO} = 0.27$, $\phi_{CH} = 0.37$) se conjugue avec un aléa influent ($\theta_{VO} = 0.49$, $\theta_{CH} = 0.52$). Ces coefficients identifiés, on peut confronter les corrélations théoriques et empiriques : sur une large plage, le modèle est fidèle. Sous des traits distincts, le voyniche et le chinois semblent trouver, sinon un parent, du moins un gène commun.

¹⁹ $S(0) = \sum \rho(\tau) = 2P+1$ compte tenu de la parité de ρ et de sa valeur en 0.



5. Conclusion

À l'aune de ses caractères, le manuscrit de Voynich ressort comme un paysage ordonné, alternant des aires denses et clairsemées par le jeu de corrélations. La résonance, sensible dans une moindre mesure pour le chinois, est inédite parmi les langues européennes, qui répartissent leurs lettres aléatoirement. Ce phénomène à mémoire longue peut se modéliser par un processus aléatoire fractionnaire.

La théorie chinoise de Stolfi est ainsi relancée. Par ailleurs, ces corrélations corroborent les observations de Montemurro et Zanette sur le plan sémantique, comme de Schinner sur le plan binaire. Visiblement, le texte en question n'est pas le fruit du hasard, mais d'une élaboration complexe. Avec la richesse et la beauté de ses illustrations, le manuscrit constitue manifestement une œuvre d'art aboutie.

Ces éléments minent l'hypothèse d'un faussaire pressé par l'argent. D'autre part, on voit difficilement un savant du Moyen Âge manier des processus formalisés au 20^e siècle. L'auteur a pu s'appuyer sur un langage naturellement rythmé pour transmettre une incantation ou une initiation. La figure de l'alchimiste s'adressant à un cercle d'initiés renaît ici, mais le manuscrit « le plus mystérieux du monde » n'a probablement pas dit son dernier mot.

Références

- 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.