Analysis on P. Oxy. 90 and Verdict

Philipp Harland

April 2025

Email: piranhafisherman@proton.me

Abstract

In this paper, we will be taking an overview of a set of results from my study of P. Oxy. 90, or Papyrus Oxyrhynchus 90, and giving a verdict as to if P. Oxy. 90 truly contains meaningful text or not.

Contents

Research Question: *Is P. Oxy. 90's mysterious cryptogram actually meaningful?*

1 Tools

Def. 1. *The **canonical structure**, $\mathfrak{C}$ of a string, S, is, informally, the "structure of repetitions". Formally, it is defined w.r.t. a string, S, as:*
$1 : \nu(l) < \nu(m)$ *if l appears before m first in S*
$2 : letter_i(S) \to \nu(letter_g(S)) \leftarrow letter_j(S), g \in \{i, j\}$ *A function, f, which satisfies 1 and 2 we call a canonical structure. However, there are infinitely many of these, and we need to add 1 more condition:*
$3 : \cup\mathfrak{C}(S) \subseteq \{x | 1 \leq x \leq length(S)\},\ \pi_1(S) = 1$ *This condition makes it unique.*
Def. 2. *The **letter frequency analysis distribution**, $\mathcal{F}$ of a corpus of text, $\mathfrak{R}$ written in a set of letters $\{\mathfrak{a}_1, ..., \mathfrak{a}_n\}$ is the function:*

$$\begin{cases} \mathfrak{a}_1 & \frac{\#\mathfrak{a}_1 \in \mathcal{F}}{size_{\mathfrak{R}}(\mathcal{F})} \\ . & \\ . & \\ . & \\ \mathfrak{a}_n & \frac{\#\mathfrak{a}_n \in \mathcal{F}}{size_{\mathfrak{R}}(\mathcal{F})} \end{cases}$$

2 Analysis and Summary of Work

The cipher couldn't have possibly been encrypted with anything more advanced than ciphers such as the scytale, Atbash or Caesar cipher – so that was our pool of ciphers for the analysis. We are also restricted to 3 languages – Greek, Latin, and Ancient Egyptian, of which the 3rd we can immediately rule out because Ancient Egyptian is abjadic, using only consonants like "k" or "t", e.x. ***km.t*** *or* ***nfrt****. with the occasional vowel in words like* ***Imnt****. The following table was generated with the frequency data of P. Oxy. 90, 91, 96, which were analyzed with a rudimentary Python program. I had picked 91 and 96 for their lengths (they could provide very helpful data), and, because they were written in the same language, same script and in a very similar time. The Python program included the base Greek letters, specifically the letter preset consisting of* $\{\varsigma, \epsilon, \rho, \tau, \upsilon, \theta, \iota, o, \pi, \alpha, \sigma, \delta, \phi, \gamma, \eta, \xi, \kappa, \lambda, \zeta, \chi, \psi, \omega, \beta, \nu, \mu\}$*, not necessarily in that order when the data in Table 1 was being computed.*
For the entries, they are numbers in $[0, 100] \cap \mathbb{Q}$*, and represent percentages out of 100.*

P. Oxy. 91	P. Oxy. 90	P. Oxy. 96
10.70	12.5	10.7
0.38	0	0.44
1.15	0	1.09
2.02	2.08	3.27
8.00	14.58	6.10
0.29	0	0.22
5.11	6.25	4.8
0.77	0	1.30
7.91	12.5	8.06
3.31	2.08	3.48
2.41	2.08	4.35
3.76	8.33	3.05
5.79	4.16	5.44
0.29	0	0.65
9.25	10.42	12.20
3.95	4.17	3.48
6.07	2.08	5.01
3.80	2.08	2.39
5.20	2.08	5.88
6.55	8.33	5.88
7	4.17	8.06
0.67	0	0
1.54	0	1.30
0.09	0	0
4.05	2.83	2.83

[1] [2] *The letters are in Greek alphabetical order, i.e.*$\alpha\beta\gamma\delta\epsilon\zeta\eta\theta\iota\kappa\lambda\mu\nu\xi o\pi\rho\sigma\varsigma\tau\upsilon\phi\chi\psi\omega$
As we see, we cannot match up the distributions without oddly permuting the domain and range of the distributions in a way inaccessible to ordinary combinations of Atbash and ROT_n.
Some analysis I had done on P. Oxy. 90's cryptogram does also not match up to that of P. Oxy. 91's text and 96's which we would expect if they were written in the same concurrent dialect of Greek in the same time. Specifically: P. Oxy. 90 does not have any consecutively repeated letters, or successive numbers in its canonical structure, which are common and even in the portion of the note separate from the 2-line cryptogram; P. Oxy. 90 has unusually long and consistent chaining repetitions, which are uncharacteristic of Greek/Latin. This is to suggest it is random and therefore incoherent; Also uncharacteristic of Latin/Greek, there are digraphs that only appear once in the entire 48 letter composition, which would very much appear in Latin/Greek blocks of similar or equal size. From what we can observe, P. Oxy. 90 has a very scatterbrained,

[1] All of the percentages have been rounded down/truncated, hence these are approximations. The raw, untrimmed data will be available in a sisterpiece to this paper.

[2] P. Oxy. 91 and 96 had been stripped of their diacritics when the data was being computed, to save some headache with programming.

spiky distribution, compared to the milder distributions of 91 and 96. It also has 5 hapax legomena, which is uncharacteristic of a typical 48 character block in 2nd century Greek or Latin, which only have 3-4 hapax legomena.[3] *We will give the following table of frequency data, generated from frequency analysis on Liber I of Lucius' Metamorphoses, a work that was made in a similar time in one of P. Oxy. 90's possible languages – Latin.*

Pct.	$\mathcal{L}$
9.14	a
1.31	b
4.18	c
2.93	d
11.30	e
1.09	f
1.08	g
0.71	h
11.25	i
–	j (omitted; 0 occurrences within the whole corpus of Liber I
–	k (omitted; 0 occurrences within the whole corpus of Liber I
3.4	l
6.20	m
5.57	n
6.12	o
2.74	p
1.37	q
6.18	r
7.35	s
8.05	t
7.91	u
1.43	v
–	w (omitted; 0 occurrences within the whole corpus of Liber
0.4754495159059474	x
0.060511756569847856	y
0 (the letter z has 0 occurrences, however it does appear in the latter portion of Liber I which was not utilized)	z

Specifically, we used the portion of Liber I of Lucius' Metamorphosis, which extended from [1] *to* [13], *from beginning of* [1] *to end of* [13]. *This should be good enough to give a useful enough frequency distribution of 2nd century Latin.*
The Latin distribution does not match up either, similar to the Greek distribution. It cannot be any combination of ROT_n *and Atbash, since the morphological*

[3]"Hapax legomenon/legomena" is only used to refer to *words*, but it saves time to use it to refer to both letters and words which appear once in a given text/corpus.

differences between the two show they are incompatible.
The earliest evidence we have for something like a homophonic cipher is from 1401, 1221-1222 years after P. Oxy. 90 was produced according to Grenfell and Hunt (it was allegedly produced in 179-180). So, it cannot have been a homophonic cipher, which is on account of the lack of evidence of such ciphers existing before that, nor is the probability sufficiently high of a 2nd century grain collector being intelligent or well educated enough in cryptography to come up with such a cipher on his/her own accord. And, since P. Oxy. 90 is so small, any attempts at a deciphering based on the hypothesis it is a homophonic cipher would be futile/too flexible.
Another line of evidence is the unicity distance. English has a unicity distance of 28 characters, which we can assume is similar to Greek's unicity distance, on account of Greek having a smaller alphabet and similar word structures to those of many English words, as well as a somewhat similar frequency distribution. 48 characters would very well surpass the unicity distance for both Greek and Latin, given Latin can also be assumed to have a similar unicity distance. However, this piece of evidence may be contested. Calculations for both will be given in a later revision of this paper.

3 Conclusion; Verdict

With all of this evidence lined up, a verdict becomes clear: P. Oxy. 90 is true to form nonsense – it is not a cryptogram, simply random gibberish with no clear direction or coherent, recognizable set of properties. As we have ruled out all 3 candidate languages – AE, Greek, and Latin – it is verifiably, nonsense.

4 Works Referenced/Utilized

Liber I, The Latin Library
Trismegistos – P. Oxy. 91
Trismegistos – P. Oxy. 96
Grenfell, Hunt, "The Oxyrhynchus papyri", pg. 152-153 (Internet Archive)