# Reviewer’s Report on Manuscript UCRY-2016-0027

## What Can Evolutionary Computation Teach Us About the Voynich Manuscript?

Personally, I very much dislike the almost cynical remarks in the introductory part, concerning early attempts of analysing the VMS (Newbold, Strong; in particular footnote 2). It may indeed be justified to see many aspects of the VMS research as ridiculous. Nonetheless, professional courtesy should dictate the usage of polite criticism in a scientific publication, rather than indulging in blatant sarcasm. After all, no researcher really is above the possibility of one day making a “pathetic error” him- or herself.

The present article is one in a series of similar attempts to identify a surmised encryption/encoding algorithm in the VMS text by “assumed text attack”. The authors’ primary hypothesis is that several “figure captions” (“labels”) placed close to the drawings of female figures in the VMS “Zodiac” section are female names of Slavic and Hebrew origin. Indeed, known text or even just knowledge about statistical properties of the plain text is sufficient to break simple ciphers like monoalphabetic substitution (the coding scheme used – at least partially – in the VMS, according to the authors).

Usually, for simple substitution ciphers the decoding transformation is reconstructed by direct application of n-gram statistics. The authors, however, utilize an evolutionary algorithm to “fit” their database containing female first names to the “Zodiac labels”. Such techniques can be extremely powerful and surprisingly efficient, but – much like the neural network approach – suffer from significant lack of control one has over its reliability. Mathematically, the evolutionary approach is closely related to numerically finding the constraint minimum of a function in high-dimensional metric space. Algorithms for both problems have the tendency to get stuck in local minima, which makes it very difficult to assess the actual confidence margin of the result.

It would be quite interesting to compare the results obtained by the authors’ evolutionary approach with those of “classical” pattern matching (on purely statistical grounds). Still, the way how a key is obtained is less relevant than its applicability. The most obvious method to validate a key is to use it for unlocking a door. The authors state that their result has to be seen as a locally optimized decoding table, where “local” obviously means, it will certainly not unlock all of the manuscript text. There are two possibilities: **(1)** their transformation table is applicable only to the labels it has originally been derived from. Or **(2)**, which appears much more likely from a cryptanalyst’s or linguist’s point of view, even a “locally limited” decoding scheme, based on a simple substitution, should be valid (to some extent) at least for a few percent of the cipher in immediate vicinity of the “crib text”.

In my opinion it is the major weakness of this work that it shows not even the beginnings of a discussion of this most crucial point. Especially since there are several strong arguments contradicting the authors’ basic hypothesis, for example:

- All evidence so far points to an origin of the VMS in the very early Italian Renaissance. This is supported by the $^{14}$C analysis and architectural characteristics in the “castle drawing”. It makes an intimate connection of the VMS text with Slavic and/or Hebrew language unlikely.
- Monoalphabetic substitution, even a (weakly) non-bijective one, does not significantly alter the word/token statistics of encrypted/encoded text. Assuming meaningful text makes finding an explanation for the (from a linguist’s viewpoint) very peculiar statistical properties of the VMS much more challenging.
- Most VMS words are written with a character set of less than 20 characters. Simple non-bijective character substitution additionally reduces this character space. Furthermore, it even worsens the problem of excessive token repetition, since also some originally different words will be mapped together.
- There is some evidence that the VMS pages were written from left to write, top to bottom.

None of these facts are set in stone. But when the authors present a hypothesis contradicting *all* of them, a significant burden of proof lies with them.

**Possibility (2):**

I would expect that a transformation valid for more than 200 labels selected from a coherent text portion (coherent statistics, same Currier language) should also “partially decode” the remaining text in immediate vicinity, at least to the extent that some linguistic structure of an underlying (Slavic) language would be more easily noticeable than it obviously is without the transformation. I applied the last entry of the authors’ table 1 (with maximal “fitness” 240) to the VMS “Bio” section (f75 to f84), since their “primary mapping” originates from a drawing in folio f84r. Apart from some Russian (on a very beginner’s level) I do not speak Slavic languages (although I should, most probably, be able to recognize some linguistic characteristics). For my point of view, the “decoded” text shows even more awkward “linguistically paradox” properties than the original VMS section. To somehow objectify this I used a state-of-the-art language identifier based on n-gram statistics that covers about 300 languages from many language (sub-) families.

The results for individual text portions of roughly 4000 characters length are listed in table 1 for both token orientations, “original” and “reversed” (as suggested by the authors). The statistics of the transformed text appears to be far from that of Slavic languages, but much closer to indigene dialects (Tzotzil is a Maya language, spoken in some parts of Mexico; Miskito is spoken in Nicaragua and Honduras).

Of course, this “identification” must by no means be taken seriously. However, I see it as demonstration of the fact that the proposed transformation indeed is driving the text towards increased incomprehensibility, rather than the opposite.

| | | Identified as Language | |
|---|---|---|---|
| **Folio** | **Characters** | **Original** | **Reversed** |
| f 75 | 4655 | Tzotzil | Miskito |
| f 76 | 5960 | Tzotzil | Miskito |
| f 77 | 4283 | Tzotzil | Garifuna |
| f 78 | 3500 | Tzotzil | Miskito |
| f 79 | 4534 | Bugisnese | Miskito |
| f 80 | 5049 | Fijian | Miskito |
| f 81 | 2767 | Tzotzil | Miskito |
| f 82 | 4028 | Tzotzil | Miskito |
| f 83 | 3998 | Tzotzil | Garifuna |
| f 84 | 4285 | Tzotzil | Miskito |

**Table 1:** “Language identification” based on n-gram statistics of VMS text transformed using one of the monoalphabetic substitutions proposed by the authors of UCRY-2016-0027 (last entry in their table 1).

**Possibility (1):**

The second remaining possibility to still save the authors’ hypothesis is much more unsatisfactory, and much more problematic: the labels close to female figures are straight forwardly encoded Slavic/Hebrew first names – but the decoding algorithm is not at all applicable outside the “figure captions” of the VMS “Astro/Bio” section. I am willing to consider this possibility just for the moment, since there are unlikely but still thinkable scenarios for such a strange situation. As an example, the encoding scheme of “labels” in the VMS could be fundamentally different from that of “regular” text.

Nevertheless, even (or perhaps, especially) such a very limited claim of successful decipherment requires substantial proof. And the only way to provide compelling evidence in such a case would be by means of rigorously proving statistical significance. Again I emphasize that the burden of proof solely lies with the authors. Still, I am going to add a few basic considerations here.

First of all, I tried to reconstruct the VMS “name-labels” list from the “Astro/Bio” section. I was able to locate 239 (including the “primary mapping”), which appears close enough to the 264 mentioned in the article. The difference should be irrelevant, and it can easily be attributed to differing opinions about which label really is to be associated with a female figure. In the following I will denote these 239 (or 264) VMS words as *names*. The *names* have an average length of 5.9 characters, which is a bit lower than the 2045/264 = 7.7 mentioned in the article (presumably, the “missing” *names* are significantly longer than the average). They are matched to a target pool of $n_{T'} = 13815$ female first names. The optimal transformation has a “fitness” of 240 characters, corresponding to 31 *names*. This means that even an *optimal* transformation will only decode about 12% of the “labels”, and nothing more.

The probability that a random string of length $L$ and alphabet size $\alpha$ will match one of $n_T$ random strings (of the same length) is given by

$$p = 1 - \left(1 - \alpha^{-L}\right)^{n_T}$$

The geometric distribution can surprisingly fast attain values close to 1 even for very low single-event probability, as long as $n_T$ becomes large enough. Nevertheless, for truly random strings the reported pattern matching would indeed have extremely significant low probability. However, the crucial point is that *neither* the source *nor* the target pool consists of random strings. Exactly the opposite:

Table 2 lists the clustering metric distance (edit distance) versus maximal cluster size for the *names* set and a set of 239 random strings (length of six characters, alphabet size = 26), respectively.

| | **Cluster Size** | |
|---|---|---|
| **Distance** | ***Names*** | **Random** |
| 1 | 1 | 1 |
| 2 | 2 | 1 |
| 3 | 28 | 1 |
| 4 | 79 | 4 |
| 5 | 164 | 58 |
| 6 | 216 | 239 |
| 7 | 232 | 239 |
| 8 | 238 | 239 |
| 9 | 238 | 239 |
| 10 | 239 | 239 |

**Table 2:** Maximal cluster size for given mutual metric distance. *Names* set and random string set with length of six characters.

For example, table 2 demonstrates that 28 of the 239 *names* are similar enough that any pair of them has an edit distance of less than or equal to 3 (characters). Now, if the target pool should be composed of strings with similar (compatible) structure, then the transformation will easily find a correspondence between them, effectively reducing the pattern matching problem from an average of 6 to just 3 random characters.

The real situation is much more complex, and only a rigorous cluster analysis could make it possible to convincingly demonstrate statistical significance. This also would have to include thorough tests with completely different sets of target strings. The authors report “micro experiments” with other data samples, but they do not report details about the statistical significance of them. In particular, why restrict the target pool to female names at all, why not using, e.g., plant names? Or random words from a dictionary of Polynesian languages? Perhaps the strange vowel/consonant chains of a Polynesian language might reach even higher “fitness” values.

Personally, I very much doubt the usefulness of such an arduous task, as well as its feasibility. And I seriously doubt the possibility that such an attempt could be successful. A first clue may be taken from figure 2 of the paper: even "misalignment" (i.e. wrong orientation, not reversed) of the strings just about halves the success probability. For a real existing "deep" correlation between the two string pools one would expect much more dramatically reduced "fitness" in case of wrong orientation.

One should also keep in mind that we are speaking about a possible decoding of 31 out of 264 labels here; 264 words out of about 7000 words the VMS is made up of. So, is there really something this special about these "labels" in the "Bio/Astro" section of the VMS – apart from the fact that they are written close to drawings of "bathing nymphs"?

Figure 1 shows the averaged "local" density of the *names* in the VMS, i.e. each point in the graph represents the probability to find one of the *names* words within the last 1000 tokens. The distinct peak at position 14000 (= folio f72) is not surprising since the labels had been extracted from this section.

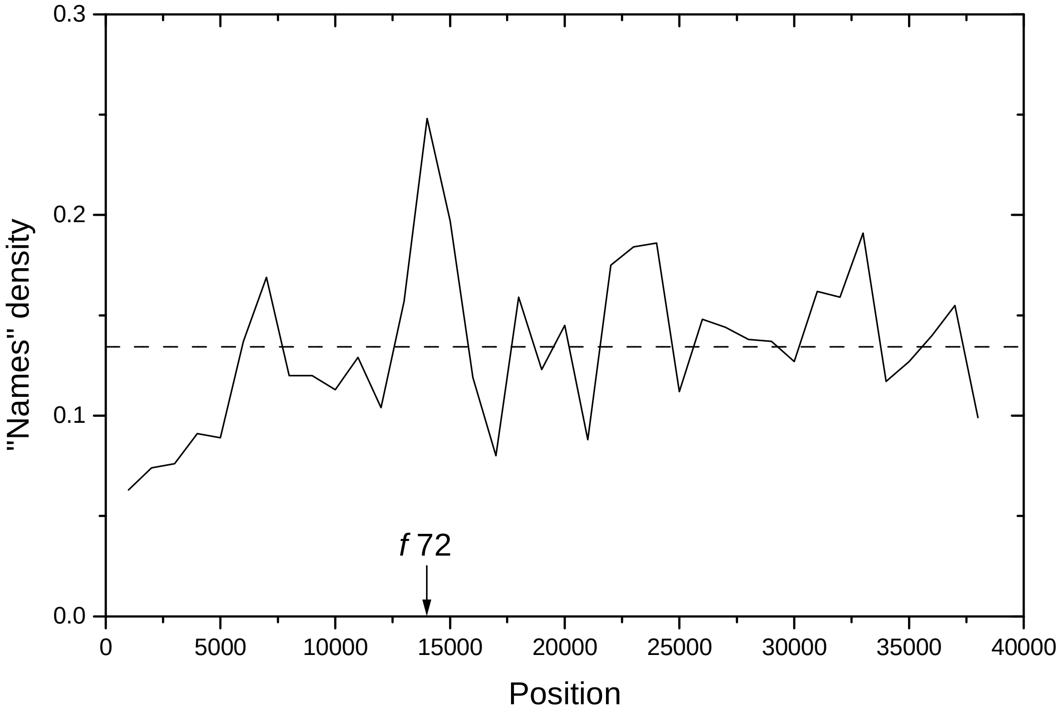

**Figure 1:** Density of the *names* (i.e. "labels") tokens in the VMS. Mesh width = 1000 tokens. The VMS tokens are numbered consecutively (= position). All of the text was used (paragraph text, labels, etc.).

But throughout the entire remainder of the VMS the density remains remarkably constant around the value 0.13. Indeed the "labels" are no "extraordinary" words within the VMS text. In fact, just the two explicitly mentioned in the paper (*okedy* and *otedy*) are among the most frequent tokens (about 0.8% of the entire VMS text). It is reasonable expecting to find “figure captions” in regular text, too. But why is then their density not more “context specific”, rather than equally distributed? And why does the proposed decoding scheme fail completely for anything else than just the *names*, despite their constant appearance in all of the text?

Summing up, from my point of view the present paper does not meet the scientific requirements for publication in a well reputed journal. The authors fail by a very large amount to present the strong arguments absolutely necessary to support their hypothesis. In fact, there even exists strong evidence contradicting it. I see no chance at all that the significant weaknesses of this work could be repaired, at least not within reasonable revision time. Therefore I recommend rejecting the manuscript.