

Numbering systems in the Voynich Manuscript

v0.1: © Nick Pelling, 8th June 2002

My core beliefs about the VMS are that:

- It dates from roughly 1480-1500 (*from the castle*)
- It doesn't include any "modern" polyalphabetic codes (*from the statistics*)
- It is written in a popular Western European language of the time (*Occam's Razor*)
- Its "code" comprises layers of cipher (*or we'd have cracked it earlier*)
- Its "cipherbet" and layout rules are built on steganography (*ie, are hidden in plain sight*)

In addition, I consider it highly probable that:

- Many of the layers of cipher/code were appropriated from pre-existing sources
- The history of the code's construction may be partially implicit in the cipherbet itself
- Much of the text matches the embedded pictures (*ie, it is herbal / pharma / astrological, etc*)
- Much of the text comprises numbers (*ie, quantities, times, angles, etc*)
- Much of the text comprises astrological values (*ie, signs, planets*)

However, apart from the [*probably much later*] foliation, and a few numbers apparently hidden in diagrams in the first quire, none of these last three is obviously apparent anywhere in the text.

Tactically, my suspicion has long been that the "weakest link" in the VMS would be **numbers**: and so it is the *numbering system (or systems)* that I have been trying hardest to crack.

The following is my current theory about how the numbering systems work - and their rationales.

(As with everything VMS-related, don't take anything I say on trust - try it yourself, see what you think.)

Numbering system #1 - measures

D'Imperio mentions Tiltman as having first suggested that the very familiar $\delta a \mathfrak{d}$ $\delta a \mathfrak{d}$ $\delta a \mathfrak{d}$ might be a numbering system (others may well have suggested this before Tiltman, I don't know). As a steganographic trick for visually hiding a sequence of Roman I's, this would make a lot of sense.

My version of this first number system is slightly extended from this. I believe that the δa portion codes for "zo" ("*oz*" backwards), with \mathfrak{z} coding for "cursive z" (hidden within a mirror image of itself).

Furthermore, in the pharma section, a number of the labels also contain $\delta a \mathfrak{z}$ which I believe codes for "zo x" (ie, "10 oncia" backwards). Note that, circa 1500, "uncia" was used as a unit of length (*inch*), weight (*ounce*), and liquid volume (*fluid ounce*), so "oz" would have been a handy abbreviation.

$\delta a \mathfrak{d}$	I oz	
$\delta a \mathfrak{d}$	II oz	
$\delta a \mathfrak{d}$	III oz	
$\delta a \mathfrak{z}$	V oz	← speculation!
$\delta a \mathfrak{z}$	X oz	← prediction!

Furthermore, the two normal VMS terminations for $\delta a \mathfrak{d}$ words are \mathfrak{d} or \mathfrak{z} . I think it likely that one of these two indicates either: "-s" (*which was an abbreviation for "-and-a-half" in manuscripts of the period*), or "-o" (*which was an abbreviation for "(prim)o", "(secund)o" ie, 1st 2nd etc*)... though I don't honestly know which is which (all the same, my suspicion is that it is \mathfrak{z} that means "-o").

I've commented on this number system before, so will leave the discussion at that for now, as you are probable more interested in the main one...

Numbering system #2 - quantities, angles, etc

This, again, has been discussed - could Voynich words like "4011c89" actually be numbers?

Certainly, they appear to have a statistical (*ie probabilistic*) distribution rather than a word-like distribution. Of course, this has led several people to comment on the possibility of the VMS being *entirely* numbers (*ie, indices into a codebook*).

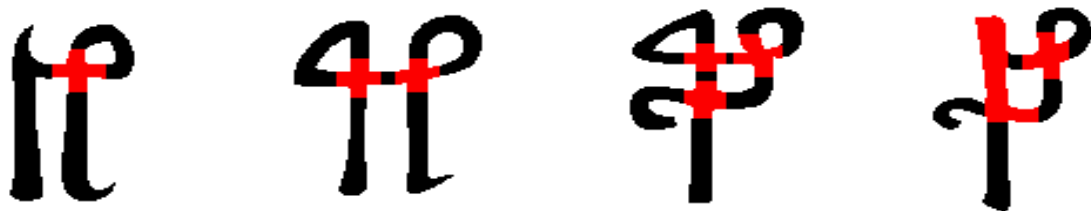
My prediction is that the VMS' main counting system runs from 1 to 98, and looks like this:-

	1	2	3	4	5	6	7	8	9	10
0+	⋈	⋈⋈	⋈⋈⋈	⋈⋈	⋈	⋈	⋈⋈	⋈⋈⋈	⋈	⋈
10+	⋈⋈	⋈⋈⋈	⋈⋈⋈⋈	⋈⋈⋈	⋈⋈	⋈⋈⋈	⋈⋈⋈⋈	⋈⋈⋈⋈⋈	⋈⋈	⋈⋈
20+	⋈⋈	⋈⋈⋈	⋈⋈⋈⋈	⋈⋈⋈	⋈⋈	⋈⋈⋈	⋈⋈⋈⋈	⋈⋈⋈⋈⋈	⋈⋈	⋈⋈
30+	⋈⋈	⋈⋈⋈	⋈⋈⋈⋈	⋈⋈⋈	⋈⋈	⋈⋈⋈	⋈⋈⋈⋈	⋈⋈⋈⋈⋈	⋈⋈	⋈⋈
40+	⋈⋈	⋈⋈⋈	⋈⋈⋈⋈	⋈⋈⋈	⋈⋈	⋈⋈⋈	⋈⋈⋈⋈	⋈⋈⋈⋈⋈	⋈⋈	⋈⋈
50+	⋈⋈	⋈⋈⋈	⋈⋈⋈⋈	⋈⋈⋈	⋈⋈	⋈⋈⋈	⋈⋈⋈⋈	⋈⋈⋈⋈⋈	⋈⋈	⋈⋈
60+	⋈⋈	⋈⋈⋈	⋈⋈⋈⋈	⋈⋈⋈	⋈⋈	⋈⋈⋈	⋈⋈⋈⋈	⋈⋈⋈⋈⋈	⋈⋈	⋈⋈
70+	⋈⋈	⋈⋈⋈	⋈⋈⋈⋈	⋈⋈⋈	⋈⋈	⋈⋈⋈	⋈⋈⋈⋈	⋈⋈⋈⋈⋈	⋈⋈	⋈⋈
80+	⋈⋈	⋈⋈⋈	⋈⋈⋈⋈	⋈⋈⋈	⋈⋈	⋈⋈⋈	⋈⋈⋈⋈	⋈⋈⋈⋈⋈	⋈⋈	⋈⋈
90+	⋈⋈	⋈⋈⋈	⋈⋈⋈⋈	⋈⋈⋈	⋈⋈	⋈⋈⋈	⋈⋈⋈⋈	⋈⋈⋈⋈⋈		

Also: note that ⋈ and ⋈ both appear in the text for "-4", so please don't give me a hard time over this - the former seems more common to me (though I could well have been misled by the encoder).

Here is my rationale for each of the component characters in turn -

- ⋈ Outside of "daiiin", this is the only character that often appears in sets of three. As such, it is the #1 candidate for "I".
- ⋈ I believe that this is steganography: rearrange the strokes to get "V". OK, this isn't the strongest argument around, but please read on...



The four gallows characters: now here's the *really* cunning bit. If you were a steganographer trying to hide the Roman numerals for X, XX, XXX and XL within a cipherbet, what would you do? The answer is obvious - you'd hide them in plain sight. If you **now** look at these characters again, what do you see? Where each loop crosses over itself, what you have is an X *suspended in mid-air*.

Clearly, this would allow us to count up to 48 without difficulty: and I believe that this is the form that an earlier version of the VMS' cipher took. However, at some point it became necessary to raise the ceiling on the number system, but without adding any substantially new characters.

I believe that the method chosen was to add a mark to the "5" character to indicate "50": and to strike through the 10/20/30/40 characters with the base of the 50 character to indicate 60/70/80/90 (which are, of course, 50 + 10/20/30/40).

Further comments: it may be that ୧ might steganographically contain a kind of "looped L" (in its extra stroke), as an aide-memoire to the encoder/decoder: and it may also be that the "teardrop" version of ୧ indicates "100" (this too would be consistent with the rest of the numbering system).

In addition, I believe that the ୩୦ prefix (*that patently often appears before numbers of this format*) indicates that the number immediately following it is an *index into a codebook*, as was frequently the case with nomenclatura and similar ciphers of the time, where the codebook would contain words that the encoder needed not to appear in the final document at any cost (like "Pappa" for the pope, etc).

Alternatively, the ୩୦ prefix could be "-s" (*ie, "-and-a-half"*), which would be consistent with its usage in Tristano Sforza's cipher (*there, it was used for "s"*): or it could be that the prefix means "number", with other numbers implicitly meaning "code-index" (*ie, the reverse*)... I don't know (yet).

For the sake of "show not tell", here are how I see the numbers on the first few lines of page f78r (one of the two pages that Leonell Strong tried to break) being constructed according to this theory:-

୩୧୧୧୧୦୧. ୧୧୧୧. ୩୩୩୧୧୧୧. ୩୩୩୧୧. ୧୧. ୩୩୩୦୧. ୩୩୩-
 20 ୧1 101 [36] [11] [10]

୩୩୩୧୧୧. ୩୩୩୧୧. ୧୧୧୧. ୩୩୧୧୧୧. ୩୩୩୧. ୩୩୩୧୧. ୧୩୩-
 [12] [11] ୧1 26 11

୩୩୩୧୧୧. ୧୧ ୩୩. ୧୩୩. ୧୧୧୧. ୩୩୩୧୧. ୩୩୩୩୩. ୩୩୩୧୧୧-
 [61] 6 10 (4?) 6 [11] [12?] 11

୩୩୩୧୧୧. ୩୩୩୩୩. ୧୩୩. ୧୧୧୧. ୩୩୩୧୩୩. ୧୧୩୩୩୧୧. ୩୩୩୧. ୧୩୩. ୩୩-
 22 [20] ୧1 [11] ୧ 71

୩୩୩୩୩. ୩୩୩୧୧. ୩୩୩୧୧. ୩୩୩୧୧. ୧୩୩. ୩୩୩୧୧. ୩୩୩୧୧. 2. ୩୩୩-
 [10] 21 [11] [11] [11] [11]

Taken as a whole, I believe that this system is elegant, clean, internally self-consistent, and may turn out to explain many of the VMS' statistical properties. But it's a starting point, not the end-point...