A recent discussion with H.R. SantaColoma on the VMS discussion list about the mathematical standard of proof needed to prove the VM is a "forgery" set me to thinking – is it possible to construct a logical inductive argument to investigate the "reality" of the VMS, and hence to say what it is, on probability, after weighing the available evidence?

Introduction

Now, I first thought it was obviously impossible to "prove" the "reality" of the VM via a purely logical argument, as we have no axioms, or truths, to base our argument upon.

However, let's reflect. Let us discard the notion of "absolute truth" or axiom, as understood in mathematics, and instead delve into the realm of abstract inductive logic, where "axiom", instead of meaning "universal truth", simply means "self obvious truth", and then deduce some postulates from which to build a logical case.

It is possible we can deduce postulates that are not immediately apparent until we consider them. For example:

- The manuscript is written upon parchment .
- A carbon dating upon a random sample of said parchment established its creation date as ranging between 1404 and 1438 with 95% probability
- etc (I expand on this below, no point in copying them all here).

There is more than enough work "out there" on the web on the Voynich. In fact, everytime I have an idea, I find that somebody else has already had that idea and has investigated it. Surely there is enough evidence out there to start getting a glimpse of the truth?

We have had over 100 years of very intelligent people looking, prodding and fiddling with the manuscript, and nobody knows what it is?

Even if nothing useful comes out of this little project, maybe it will help direct further research effort / or just be useful for the links within.

I must stress one thing. I started this project with an entirely open mind, or as close to one as possible. I simply *didn't know what the Voynich could be*. It baffled and intrigued me.

So I started to investigate everything I could. I spent a month amassing information and reading before starting to put all this together. At the end, I had a conclusion. Which is the point of this exercise. Your final conclusion may be different. Who knows the truth?

David Jackson

August 2014

www.davidjackson.info/voynich

voynich@davidjackson.info

Version 1. last update: 19/08/2014.

Table of Contents

The Hypothesis

I start by defining a hypothesis to prove, and a null (contrary) hypothesis.

- H0 : The Voynich Manuscript contains understandable content.
- H1 : The Voynich Manuscript does not contain understandable content.

I decide upon a progressive logical argument.

I collect all of the most common "explanations" for the Voynich Manuscript, collate all possible third party analysis of each explanation, and attempt to evaluate these analysis.

(I have discarded all theories that are backed by only one person or source. There are so many theories on the Voynich "out there" that some filtering mechanism is necessary. I attempt to use crowd control as the filter – if the internet hive mind does not seem to accept a theory, I do not consider it. It is a primitive form of peer review, if the theory of origin does not pass, I discard.)

The evaluation consists in a logical deduction from the available body of I evidence that I collected which I term a *Conclusion.* Each conclusion is numbered sequentially, starting at C1.

Definitions

Let us start by defining terms.

- **VM** (or VMs): Voynich Manuscript. The original book or text.
- **Forgery**: I take "forgery" to mean the book was made in later years with a deliberate attempt to pass itself off as a 15th century text for financial gain.
- **Hoax**: I take "hoax" to mean the book was made in later years with a deliberate attempt to pass itself off as a 15th century text, but not for financial gain (ie, an art project), but still with intent to deceive.
- **Experts**: The VM field is one that has been neglected by traditional academia. This does not mean that people from many walks of life have not spent considerable time examining the VM. Therefore, when I use the term "expert" I will attempt to reference in which field said expertise lies.
- **Theory of origin**: A theory about what the VM "is". Ie, ancient herbal, Mayan text, fountain of eternal youth donated by extraterrestrials, etc.
- **Theory of generation**: A theory about how the VM was generated. Ie, encoding / decoding attempts, mathematical analysis, etc.

Origins – Materials

Let us take René Zandbergen's "Origin of the manuscript" page as a reference, and the report of the analysis of the ink. René writes:

> "In 2009, the University of Arizona performed a radiocarbon dating of the parchment. Four samples from different locations in the MS were individually tested, and the probability curves of their C-14 content were fully consistent with each other. The combined date of the parchment was established as ranging between 1404 and 1438 ACE with 95% probability.
>
> At the same time, the McCrone institute analysed the inks and the pigments used for the writing and painting of the MS respectively. A detailed report of this analysis has been made available (here), and to summarise here, it was stated that there were only minor variations of the ink composition throughout the manuscript, and iron gall ink was used everywhere. The pigments were based on ground minerals of high quality, for which recipes existed since the middle ages. There was no sign of any constituent which was inconsistent with the radiocarbon dating of the parchment. The pigments were also considered to have been moderately expensive. "

I would add that the report concludes that there is a probability that the ink used for the quire numbering, the page numbering and the ink used to write the Latin alphabet differ from the text/drawing ink and from one another.

The vellum binding of the book was analysed by an expert in such matters (De Ricci, Seymour: with the assistance of W. J. Wilson: *Census of medieval and Renaissance Manuscripts in the United States and Canada*, 1937.) who concluded that it was added in the 16th century, during a rebinding of the manuscript. This was a normal occurrence at the time. The limp vellum used to bind the outside of the VM was of decent but not excessive quality.

There are, including blank pages and pages with no text, a total of 240 pages. 225 pages include text.

Let us summarise the conclusions and draw up some new postulates, which I call Conclusions:

- C1: The parchment of the manuscript is from the first half of the 15th century.

- C2: The ink used on the parchment was available at this period in time.

- C3: The ink used for the writing and the drawings is essentially the same.

- C4: The ink used for the numbering of the quires / pages, and the Latin alphabet, differ from one another and the writing / drawing.

Origins – the inking

Here I look only at the *inking* of the manuscript, not the alphabet or glyphs.

The inking of the VM is generally divided into the following sections: The main body of text, the illustrations, the page numbering, the marginalia and the "chicken scratches".

The debate about how the inking of the VM was carried out continues to rage. However, there are a number of detailed analysis out there which I have attempted to collate. I leave the question of whether or not there were two scribes (the Currier A&B argument) for a later section, as I know of no arguments that the style of inking was different between them (other than the handwriting style).

First, **the main body of text and illustrations**:

[René Zandbergen](#):

> *Looking at the production of each single bifolio (or foldout folio), there is little or no doubt, that the illustration outline was drawn first. Secondly, the text was written, carefully avoiding the illustation. There are, however, cases, where the text interleaves with the illustration, and the size of the gap between (for example) braches of a herb exactly fits the size of the word written in between. This has been seen by some as an indication that the text is meaningless filler, but it could also mean that the page has been copied from a well-prepared draft. Whether all four drawings on a bifolio were done before all four text sections, or it was done page by page, seems impossible to say. The painting of the illustrations was done afterwards.*

Reddy & Knight, [*What we know about the Voynich Manuscript*](#): (page 2)

> *The text was probably added after the illustrations and shows no evidence of scratchings or corrections.*

D'Imperio in [*An elegent Engima*](#) also concurs (page 11) when she quotes Elizabeth Friedman and Dr A.H. Carter who come to separate but identical conclusions that the same person wrote and drew the illustrations at the same time.

There are many more independent analysis out there that come to the same conclusions. The illustrations were sketched and then the main body of text inked around the illustrations, as would be logical.

I come to the conclusion that:

- C5: The illustrations were sketched first and then the text added afterwards.

Caveat to C5: There is reasonable doubt about whether the "colouring in" of the illustrations was done at the time of sketching the outlines or later. I do not go into this argument at this time, they are long and arduous.

Quire / page numbers.

Nick Pelling has an excellent treatise of why he believes the book was bound out of order, and the quire / page numbers added later which can be read here. I know of no major arguments against this.

I come to the conclusion that:

- C6: The quire and page numbers do not correspond to the true pagination of the VM as indicated by the illustrations and text flow.

From this conclusion, we can consider two reasons for this discrepancy:

1. The numbers were added by someone who was not aware of the true pagination of the VM.
2. The numbers were added by someone who was aware of the true pagination of the VM but had a reason for obfuscating it.

Since it appears (we do not conclude as there is no evidence other than supposition) that the binding was done at a later period, and Conclusion 4 says the ink is different, reason (1) seems more likely. An attempt to order the pages at the time of a re-binding could well involve the binder marking the pages in the order he thought they should be.

I therefore suggest we discard the numbering from our analysis as not relevant, and probably as being added later by an unknown owner of the VM.

Marginalia and the chicken scratches

By their very nature, they do not form part of the main body of work and there appear to be no theories linking them to the original work. In either case, they are so indistinct that debate rages about their very nature. Therefore, I suggest we discard them as irrelevant to this study. They were probably added later in the history of the VM.

Origins – the glyphs

Debate continues about the very nature of the alphabet employed in the VM. How many different glyphs are there? Is there punctuation? Are there vowels and consonants? There are so many questions.

Let us break this down into the main questions that are being asked time and again, and see if we can work out some basic conclusions.

How many glyphs are in the alphabet?

Nobody knows. Stolfi identifies 7 different transcription alphabets, plus EVA. Each differ.

Why is this?

Partially, this is because there is a lack of text. With only around 35,000 "words" (as defined by the space between glyph blocks) there isn't enough text to be *unambiguous* about it.

Another major problem is the style of inking the glyphs. As is common with ink quills, the glyphs are made up of a series of pen brushes, and often flow one into another (joined up writing). Therefore, exact agreement on where a glyph ends and starts is lacking. The glyphs are almost fractal like in nature – every time you take a closer look you think there are more glyphs there (as taken to extremes by Newbold's stroke theory).

However, transcribers generally settle upon alphabets of between 20 and 30 glyphs, with which they can account for virtually all of the text. Additional glyphs that appear very rarely are generally called "weirdos" and although transcription efforts attempt to account for them, privately they are often dismissed as scripting errors by the author.

The issue with this is that it becomes a major issue for anyone attempting computational attacks on the VM – he or she has to choose the transcript they most like the look of, and hope to hell they've got the right one :)

In fact, there isn't even a consensus on whether an alphabet *exists*. There are many theories out there: it's shorthand, words are syllables, etc.

However, the fact that transcribers *do* generally settle upon a set range of glyphs indicates that there is an underlying alphabet.

This leads to a very unsettling conclusion:

- C7: There is no consensus on the Voynich alphabet.

However, this is not necessarily due to a fault in the VM. In order to machine compute the text, we 21st century people must first transcribe the VM into our Latin alphabet, without understanding the base alphabet. Rather a chicken and egg situation.

Let us assume we are the author of the VM. We have our base alphabet, but must write this with a quill in a series of strokes. We have no proof reader (for there are no scribal corrections to the text)

and so we merrily start writing. Round and up and over the gallows. Sometimes, to use a modern equivalent, our handwritten *a* comes out as an *e* or a *u*. Or maybe we make a mistake and insert a character that doesn´t really exist, such as a œ. A mistake, but a harmless one we think to ourselves. Except that 500 years later, somebody comes across the œ and thinks "what the blazes is that?". Some transcribe it as o e, others as œ.

How secure are we in our transcripts?

For the very reason in C7, any attempt to use transcripts in analysing the text undermines our analysis. "Garbage in, garbage out" as they say. So any computation attack on the VM must attempt to minimise these errors by being re–run on all available transcripts, and an average being worked out from the different results. EVA, for example, is a stroke transcription rather than a glyph transcription – many glyphs are transcribed in two or three strokes. So it's not that much use for statistical analysis unless this fact is taken into account.

- Corollary to C7: Analysis using only transcripts must attempt to adjust for erroneous input.

Is it a natural language?

This is the wrong question to be considering, although it is always the first to jump out.

By "natural language", most people assume a language like English, Spanish, Chinese or Arabic. A standard set of "words" that everyone understands, that map sounds (speech) and ideas to ink on paper.

But English, Spanish, Chinese or Arabic all differ between themselves. They are all "natural languages", but English has a fairly arbitrary mapping of characters to pronunciations, whilst Spanish has a strict one. Chinese is character based, it doesn't use letters but symbols. Arabic is an abjad, a series of consonants and the reader must mentally insert the vowels as they go along. Four very different ways of representing sounds.

This is the first step towards gaining any insight into the linguistic arguments that surrounds the VM. Let's break the "natural language" question into more digestible sub questions.

Is it any known writing system?

No. The VM alphabet is unknown. Nor does there appear to be any one to one connection with any known writing system of the era. D'Imperio (1978) gave an exhaustive review of the topic and concluded that there were no known manuscripts comparable at that time; none have been found since, although occasionally an individual glyph is found that bears a similarity to one in another manuscript. These "discoveries" are usually dismissed as coincidence.

- C8: The VM is not written in a known writing system.

Is it a type of shorthand?

Tironean notes were the shorthand of the Middle Ages for the scribe. The legal scribes of Castilla, for example, went mad on them from the 13th century onwards, creating a confusing and difficult code to speed up note taking and official documents in the courts. It is perfectly possible that if the VM were a type of personal document, Tironean notes would have been used.

However, the difficulty with this is that analysing the words doesn't give any clue to an underlying language. Normally, Tironean notes simply encode *part* of a word, for example, standard Latin endings such as –us. Reading a page written in the style still quickly lets you know that a) it's in a specific language and b) some shorthand is going on.

Another specific problem with this theory is that the labels don't indicate this. Single words (as the labels are) should be direct and to the point (it is logical to assume they are naming the drawing they accompany). The single labels are as inscrutable as the main body of text.

It is possible that the language is encoded or encrypted in some fashion, and an additional layer of shorthand is added to reduce writing time. This is unknown at this time, but should be born in mind if anyone is attempting to discover a code in the VM.

- C9: The VM is not written in shorthand alone.

Is the VM an invented ("constructed") language or based on an existing natural one?

If it is a constructed language, it is of a different level to any other known invented language of before the 19th century.

Invented languages were known during the Renaissance, and indeed, discussed and developed.

On the "crackpot" side we have such things as Enochian by John Dee in the 16th century. Despite having an invented alphabet, linguists believe this to be a sort of "glossolalia", invented babble committed to paper. Many words have a strong similarity to English or Latin. It's generally thought that the whole thing was a scam. Another earlier example is the Lingua Ignota, described by the 12th century abbess of Rupertsberg, Hildegard of Bingen, who apparently used it for mystical purposes. To write it, she used an alphabet of 23 letters, the *litterae ignotae*. She mixed in words of her own devising into standard Latin, possibly to create a secret language between initiates. Ramon Lull in modern day Catalonia would produce a logical language a half century later, in which his *ars combinatoria* , a series of concentric paper circles, allowed a user to input a question, have it translated into a "logical universal language" and a solution prepared. Some claim that this is the birth of modern Logic and computer processing, but they would.

More serious works from the 16th century onwards include Balaibalan from the Middle East in around the 16th century. Balaibalan is an a priori language, written with the Ottoman variant of the Arabic alphabet. The grammar follows the lead of Persian, Turkish and Arabic; like Turkish, it is agglutinating. Much of the lexis appears wholly invented, but some words are borrowed from Arabic and the other source languages, and others can be traced back to words of the source languages in an indirect manner, via Sufi metaphor.

Dante (of the seven circles of hell fame) was another one who enjoyed making a "universal language", *de vulgari eloquentia.*

There are many more. Wikipedia has many "constructed languages".

However, when I look at these languages I see three main types:

1. Glossolalia. Nonsensical writings such as Dee's, which do not exhibit any regular patterns.
2. Adaptation of existing languages such as Hildegard's, which are either attempts to obfuscate or clear up the meaning . The bulk of the text is a well known language with modifications – either cipher–like substitutions with invented words, or attempts to make the language more accessible (such as the many 19th century attempts to create a phonetics English).
3. Entirely made up languages which had an intention to be used. In which case they had a clearly defined grammar, were usually based on some system that people could be assumed to know already (ie the usual alphabet) and tried to be simple.

The VM, it is clear, would fall into category (3), if it were an invented language, and would be considered to be a fully *a priori* constructed language.

We know, as mentioned above, of several constructed languages from the 15ᵗʰ century onwards. However, in the European tradition, the earliest constructed languages had a strong religious

connotation, such as Dante's Christian attempt for an ideal Italian for literature or the Kabbalistic attempts to recreate the language of Adam and Eve. The VM does not seem to contain any known religious imagery (see *concepts*). I am unaware of any attempts from this era to create a "universal language" that did not include these concepts.

Later Renaissance attempts to create constructed languages often revolved around alchemical or magical concepts. This would appear to fit in better with the imagery found in the VM. Constructed languages throughout history have tended towards the ordered and structured. Often, indeed, even as far back as Ramon Lull's 13th century logical language, words and sentences are successively repeated and differ only on the substantive, the adjective, the verb of the proposal or on another symbol used as a changer of reference (Umberto Eco in *The search for the perfect language in the European Culture*).

Several independent mathematical analysis of the VM language again appear to preclude the possibility of it being a natural language, as they have all detected a strong ordering to the layering of glyphs within words that are inconsistent with a language. The underlying construct of the language is simply too ordered for it to be a naturally generated language.

There are almost no repetitive phrases in the VM. Notwithstanding, there are a number of very regular rules in the creation of words and sentences. We can describe adjacency rules on a letter basis within words. There are letter patterns that occur a lot – others that never appear. In a text based on natural language, the standard rules of grammar should exist between words, these relations causing the words to appear in the same order multiple times.

For example, Antoine Casanova (2000), *Rules in the Voynich Manuscript,* describes a synthetic rule building mechanism for Voynich words from which he concludes:

> "The terms of the Voynich manuscript are built from synthetic rules which exclude the assumption from the use of a natural language for its writing. However, the rules which we have put forward could be the expression of a progressive modification, inspired from the discs of Alberti, from the encryption used by the writer(s) of the manuscript."

Jorge Stolfi later came to much the same conclusion with his "crust-mantle-core" paradigm, from which he concludes:

> "The nature and complexity of the paradigm, and its fairly uniform fit over all sections of the manuscript (including the labels on illustrations), are further evidence that the text has significant contents of some sort. Moreover, the paradigm imposes severe constraints on possible "decipherment" theories. In particular, it seems highly unlikely that the text is a Vigenère-style cipher, or was generated by a random process, or is a simple transliteration of an Indo-European language. On the other hand, the paradigm seems compatible with a codebook-based cipher (like Kircher's universal language), an invented language with systematic lexicon (like Dalgarno's), or a non-European language with largely monosyllabic words."

Both models allow for prediction of the frequency of words and the construction of such words with considerable accuracy.

However, Stolfi has flagged up [a very important fact](#) which has so far passed unnoticed by many: the length of the words is exactly distributed around the arithmetical mean with the average word length being 5,5 characters. This is very unusual and indicates a mathematical generator.

A constructed language would not display this characteristic unless the constructed words were of a unified length, something that wouldn't necessarily have made sense in the pre-computer era. It would make sense if the words were being generated in some fashion (either by an encryption method or a random word generator) as the method used to produce the words would have a simple generator.

Torstein Timm in *How the Voynich Manuscript was created* identifies numerous examples of similar words being used in distinct word orders. For example, he identifies:

> The text of the VM seems to be unique because repetitive phrases are missing. On the one hand, it is possible to describe adjacency rules on the letter level. There are, for instance, typical letter combinations but also letter combinations that never occur. On the other hand, there are no similar rules on the word level. Only a few repetitive phrases can be found. There are only 35 word sequences, which use at least three words and appear at least three times.[5] Only for five of these sequences is the word order unchanged for the whole manuscript, whereas for 30 out of 35 phrases the word order does change.[6] Does this mean, that it is simply coincidence that these five sequences did not appear in a different order? An additional observation is that in 24 out of 35 cases these repeated sequences use at least two words which are either spelled the same or very similarly. Does that mean that they are only repeated because they use similar words? For instance, the words 1oe ("chol")[7], 2oe ("shol") and Koe ("cthol") occur together three times.[8] Each time, the word order is different:

<f1v.P.6> *[glyphs]* ("chol cthol shol")

<f4r.P.2> *[glyphs]* ("chol shol cthol")

<f42r.P2.10> *[glyphs]* ("cthol chol shol")

He continues to identify a whole series of words grouped by their similarity and frequency. However, in 2014 a new theory appeared which veers the argument back towards the possibility of a "natural language", after a lapse of about 30 years since any previous such theory.

Professor Stephen Bax has published a substantial treatise arguing that [the VM is a natural language written in an unknown script](#) and has proposed translations for a number of words. [Nick Pelling isn't convinced](#) although Professor Bax has [published a rebuttal](#). However, Professor Bax's translation is limited to single words and he has so far not published any work expanding the translation to entire sentences or understandable text. He also has not published any grammar or further explanation on the work or adequately explained away the features noted above. He *does* repeat the very good point

I made in C7 – if the transcripts do not adequately reflect the original intent of the scribe, what good are they?

A further concept to explore is whether the script is an abjad, a script that omits vowels and only writes down the consonants. The first to published a major work on the subject was John Stojko, who published "Letters to God's eye: The Voynich manuscript for the first time deciphered and translated into English", claiming it was written in "voweless proto Ukrainian".   Joachim Dathe more recently thinks it's poorly written Arabic. Reddy & Knight apply a two state bigram Hidden Markov Model to the EVA transcription and finds evidence it is an abjad, namely that the word grammar is a*b, but this appears, in my personal opinion, to be a mis-use of the EVA transcription and a subsequent misinterpretation of the above results.

However, let us consider the possibility that the VM is an abjad. How can we refute this theory?

To detect an abjad, we'd have to discover a Consonant Vowel (CV) structure in the text. That's the 2 state HMM mentioned by Reddy & Knight. Basically they want to find repeating CVCV blocks in the text. Nobody has ever found any, that I'm aware of.

In fact, the VM drifts between the two. It has an internal structure similar to a full alphabet, yet presents elements of an adjab. But before discussing this, we have to decide which glyphs are vowels and which are consonants – and nobody has yet managed to find enough evidence to do this.

Now, an interesting concept that has been explored are codes. Is the VM encoded? It is possible. I discuss this next.

From the above, we can conclude that the latest evidence is telling us that:

- The VM is missing any structure that would indicate the usage of a natural written language which uses consonants and vowels.
- The VM may be exhibiting elements of a phonetic language (*ie Chinese*), but the lack of repetition throughout the text, and the mathematical construct of the words, do not appear in such phonetic languages.
- Because of the above, the VM is not a direct cipher. It does not encode "word for word".
- We are unsure which glyphs are vowels and which are consonants.
- The concept of a "constructed language" was in existence before the VM was produced.
- The VM appears to have a mathematical generator for its words which cause them to have a large arithmetical mean.
- VM words are easy to predict – they have strong internal glyph order.
- The above all depends on how good our transcriptions are.

From the above, I conclude that:

- C10: The VM is not a natural language.
- C11: The VM may be a restricted constructed language.
- C12: The VM appears (remember C7!) to have a strong underlying pattern that hints at a generator.

Is the VM in code?

The mathematical analysis, as described above, hints that the VM is in code.

However, nobody has even come close to finding a code that would fit the findings above. But let us consider the possibility.

Before delving into the depths of Bletchley code breaking, let's stop and consider reality.

The VM is fluently written. The scribe was writing in a practised, fluent and confident manner. Although some people have found evidence of minor alteration to glyphs nobody has ever discovered evidence of corrections. The scribe was not carefully copying another text he was unsure about, he was speeding along and writing something. If he was stopping and working out the cipher as he went along, it can not have been a very laborious process. From the way the ink is flowed onto the page, it seems the scribe would write several words in a row before dipping his pen into the inkwell to refill and continue.

I conclude that:

- C13: The VM was written in a fluent and confident manner.

Also, let us consider the provenance of the VM. We have concluded that materials are from the early 15th century. It is entirely possible, if not probable, that the VM was produced in that time frame. So we are dealing with a cipher that must be easy to produce and read.

There were no computers. Any cipher or code that has ever been found from that time is limited in its nature *to the essential.* We can develop great theories about quantum encoding processes, but this is irrelevant to the VM. Any cipher would be limited to a) the realities of the era's philosophical and mathematical abilities and b) the realities of the encoding / decoding process.

Now, I concede that C7 may be limiting our ability to machine attack the VM. By limiting our study of the VM to the transcriptions, we are ignoring many possibilities, such as characteristics of the glyphs, visual encoding clues, and the like.

Let's consider some ciphers and codes.

The simplest ones can be discarded. Caesar ciphers or simple substitution ciphers have been explored *ad nauseum* with no success. And they don't produce what we're seeing with the VM language. Which is a pity, because they are the obvious answers: they were known at the time, and they can be learnt in an afternoon and applied visually as you go along.

Nor could it be a code. In a code, a symbol stands for a concept. A "code book" is used to translate the symbols into their meanings. The way the VM is constructed means that there would be tens of thousands of potential codes, meaning the "code book" would be considerably larger than the VM itself!

Codes from the era, especially diplomatic ones, were obvious. Two people, two code books, a bunch of pre-arranged phrases indicated by symbols. The Kings of Spain loved them and insisted upon using them. Look at diplomatic texts from the time, and you'd have introductions, then a code symbol instead of a name, some text that there were no codes for, then a whole bunch of symbols, and finally "in the name of God and your Gracious Majesty, etc etc".

And to use such a code, you have to keep stopping and encoding the next bit of your message. You don't fluently write it as you go along.

- C14: The VM is not a simple substitution cipher nor code.

What about a "sparse" cipher? The trick here is to enclose your message in random noise. The traditional one, and the first to be described during the first half of the 16th century, was the "Trithemius cipher". Make a tabula recta, a square table of alphabets, each row of which is made by shifting the previous one to the left. Data is encrypted by switching each letter of the message with the letter directly below, using the first shifted alphabet. The next letter is switched by using the second shifted alphabet, and this continues until you have encrypted the entire message.

Trouble is, these things are much harder to use for long messages than you'd realise. People have experimented time and again and found nothing. And test texts encoded in this fashion don't produce anything like the VM text.

And besides which, why wrap up this cipher in a strange alphabet? It's not only against all logic, it's unheard of in history.

No, to create a cipher that produces Voynich *like* text you need to use concepts and methods from the modern era. At a minimum, we *know* the VM was from before its public discovery in 1912, so it *must* use concepts from before that era.

We can postulate that it uses multiple layer encryption; we that the plain text was broken apart into syllables or CaMeLcAsE or phonetic groupings before different layers of encryption were applied, but a very very simple fact remains:

If you were bright enough to work out a cypher this complicated in the Middle Ages, and had secrets enough to hide, why the hell invent some funny alphabet that's only going to attract attention?

No, I'm afraid that any attempts to continue to insist that a multiple layer encryption has been applied to the text will only fool ourselves.

- C15: The VM is not written in a sophisticated cipher nor code.

Is it Gibberish?

Sadly, it's much easier to say the text is Gibberish than to say it's a code.

Gordon Rugg was perhaps the first to describe how to generate the VM by using random text. He described his Cardian Grille system to generate text, but flaws in his system have been discovered. For example, the generated text does not correspond to the statistical properties of the VM.

Rugg was able to modify his Grille system to fix these flaws, but only by adding in an unnecessary level of complexity into his system.

But Torstein Timm in his academic paper "How the Voynich Manuscript was created" describes a very simple process for creating the Voynich Manuscript using 15th century methods. He identifies the fact that similarly spelt words appear close to one another in the text and builds a "grid" of VM words, defining relationships between them. He then realised that most related words in the VM appear close by one another on subsequent pages.

His theory is that the text was "generated" during writing by copying earlier words and modifying glyphs as the scribe went along, in order to produce nonsense text.

He identifies an issue with his theory in the "labels". In essence, he shows that there are pages where words of a particular series are frequently used, and pages where they are rare (page 8 of his treatise).

He also identifies labels in different sections that are the same, and asks why a "label" attached to a star is the same as the one being used in the herbal section – why would a star be called the same as a plant?

A possibility is my "Volvelle" theory, in which a wheel is used to start the generation of words. A series of concentric paper rings are constructed, with different glyphs on each section. The wheels are rotated to give new words. This accounts for Stolfi's "crust core mantle" paradigm as glyphs from each section are simply the ones on each ring.

The scribe uses the Volvelle to generate the first word of a sentence. He then uses Timm's system to generate the rest of the text as he goes along. For labels on pages with many similar words he uses Timm's system – on pages where there are no continuation between concepts he uses the generator each time and generates similar looking words (perhaps the scribe only moves the middle ring or the last one each time).

- C16: A simple and fast process to generate random text for the VM was available.

Conclusions

User-Defined Index

Looking at the above list of conclusions and applying to our Hypothesis test, what do we find?
Well, it is not written in a known writing system. It is not written in shorthand.
It is not a natural language (although it may be a restricted constructed language). Although it has a strong underlying pattern that hints at a generator, it was written in a fluent and confident manner. It is not a cipher nor a code, and a simple and fast process to generate random text for the VM was available.

I would say, having weighed over 100 years worth of study of the VM in my hand, that we have three possibilities:
- The VM is written in a long lost script and language. In which case, unless we find more examples or a key, we'll never know what it says.
- The VM is written in an amazingly sophisticated code. In which case, it's probably a modern day hoax.
- The VM is gibberish.

So, the VM does null hypothesis is to be favoured:
the VM **does not contain understandable content**.

Postscript.

Bugger.

As I said at the beginning (now written several weeks ago), I had no axe to grind or pet theory when I started this. I simply wanted to "weigh the evidence" that was out there.

Along the way, I ran loads of experiments to check the veracity of other peoples work. All seemed to be on the level.

But it was Torstein Timm's analysis that got under my skin. So I started looking at the theory, and then came up with the Volvelle theory (described on my website) and the world sank away as I realised the whole edifice was hollow.

In fact, it has only just sunk in now.

I have a theory up here which continues this line of thought. Since it's speculation, I've removed it from this document.

What now? Well, off to the pool I reckon. And a nice glass of *vino.*