An Anatomy of Failure:

Investigating the Reasons behind Failed Attempts to Decrypt the Voynich Manuscript through an Examination of the Decipherment of Linear B and of the Partial Decryption of Kryptos

By Adam Lewis
Acknowledgements:

Many thanks to George Erving, Tony Mullen, Adam Smith, Ti Allshouse, my friends in the Honors program, Sara, Deanne, and countless others without whom I couldn’t have gotten this far.
Table of Contents

Table of Contents.................................................................................................................3

I. Abstract.........................................................................................................................4

II. Introduction....................................................................................................................5

1. What is the Voynich Manuscript?.................................................................................5

III. Features of the Voynich Manuscript..........................................................................5

IV. Other Codes................................................................................................................8

1. The Decipherment of Linear B.....................................................................................8

2. Kryptos........................................................................................................................13

V. Investigations of the Manuscript.................................................................................18

1. The Hoax Hypothesis.................................................................................................18

2. An Amateur Attempt..................................................................................................23

VI. Conclusion..................................................................................................................25

VII. Bibliography..............................................................................................................27

VIII. End Notes................................................................................................................29
Abstract

The Voynich Manuscript is a medieval manuscript whose creation dates back to the 15th century. The manuscript is written in an unknown script, and since its discovery there have been no successful attempts to decrypt it. This paper notes the peculiarities of the Voynich Manuscript and compares it cryptographically with Linear B and the CIA’s Kryptos monument. This paper examines the decipherment of Linear B and the partial decryption of the CIA’s Kryptos monument in order to discern what elements made those decryptions successful. There is not a large enough corpus of “Voynichese” text to perform any significant breakthroughs in translation, and there is evidence supporting the hypothesis that the Manuscript may be a hoax with no actual meaning contained within. Although all decryption attempts were unsuccessful, only some were unsuccessful for cryptographically sound reasons, while others never had a serious chance. The three elements necessary for a successful decryption attempt are a large enough corpus size, an external reference point for verification, and a methodological approach which consists of making reasonable hypotheses firmly grounded in data.
Introduction

What is the Voynich Manuscript?

The modern origins of the Voynich Manuscript lie in 1912 when it was discovered in the private collection of book collector Wilfrid Michael Voynich. Voynich found it among a collection of illuminated manuscripts he had purchased, however how the manuscript came to be in this collection is not known for certain. Upon its discovery, the manuscript was originally thought to be simply a collection of exotic plants, with text written in a substitution cipher, and that a botanist and mathematician working together could easily decipher it. This was not the case. Over the years many theories have gone around regarding its content and its authenticity, but nothing is known for certain. In 2009, scientists at the University of Arizona radio carbon dated the manuscript with 95% certainty that it was created sometime between 1404 - 1435, which adds credence to its authenticity.

Features of the Voynich Manuscript

The most noteworthy feature of the manuscript is that it is written in a totally unknown script. No other examples of the Voynichese language, as it is called, exist. The other most obvious feature is the collection of strange pictures contained within, such as nonexistent hybrid plants and nude women bathing in bodies of water shaped like internal bodily organs.

The manuscript is commonly divided into five to six different sections, based on the nature of the illustrations in each (Montemurro 1). The sections are: “Herbal,” “Astronomical” or “Cosmological,” “Biological,” “Pharmacological,” and “Recipes” or “Stars.” The Herbal section is the longest of the sections in the manuscript, and is so named for the pictures of false
plants and plant composites it contains. Depictions of “plant composites” are plants which have leaves of one kind of real plant, stems of another, and roots of a third, for instance. The “Astronomical” or “Cosmological” section is so named for the illustrations of the Zodiac and other heavenly bodies contained within. The “Biological” section is so named for the illustrations of nude women and bodies of water shaped like internal organs. The “Pharmacological” section is named for the jars (thought to be medicine jars) illustrated within. The “Recipes” or “Stars” section contains small stars drawn in the margins, and is hypothesized by some to potentially be a list of recipes.

The manuscript is approximately 240 pages long, of which some are missing and some are not in the correct order. The Voynichese script can be divided into two different “dialects,” designated “Voynich A” and “Voynich B.” The two dialects differ in letter frequency, word length, handwriting, and what section of the manuscript they appear in (Rugg 34). For instance, a large part of the Herbal section is written in “Voynich B” but other sections are written in “Voynich A.” There are somewhere between 20-30 different characters (Timm, 2016). Voynichese words appear to consist of a prefix-infix-suffix structure (Rugg 34-36).

Voynichese exhibits an interesting statistical property: the length of words in the script follow a binomial distribution along the mean (Timm, 2016). That is, most of the words in this language are closer in length to the overall average word length (about 5.5 letters) than there are short or long words. In contrast, English (which has an average word length of 5.1 letters) has a flatter distribution, with more short words like “a”, “an”, “of”, and “the”, and more long words like “luminescent”, “circumlocution”, “idiosyncratic”, and “incomprehensibilities”. Voynichese
is interesting in that a word’s position in its line on the page tends to be a predictor of its length (Timm 2016 1-2). Longer words tend to appear near the start of lines, and shorter words tend to appear near the end of lines. Vöynichese has no discernible punctuation markings, so a single line of text appears to be roughly analogous to an English sentence (Rugg 34). We can suppose this because of the word length pattern as well as due to other patterns. One such pattern is the appearance of so-called “gallows letters” that tend to appear at the start of lines. Gallows letters are more ornate than most other letters in Vöynichese; they are taller and have looped sections at the top of each character. Likewise, there are also certain characters that are more likely to appear nearer to the ends of lines.

Another set of interesting properties in the Vöynichese text is the phenomenon of context dependency and word co-occurrence. Context dependency refers to the case in which words with similar morphological structure tend to cluster and appear very near each other in the text, and word co-occurrence refers to words repeating multiple times in sequence, roughly equivalent to the English “had had” or “the the the”. Both features are unusual among languages. One researcher, Torsten Timm, describes the phenomenon of context dependency:

In natural languages a word normally (cf. poems) is used because of its meaning and not because it is similar to a previously written one. The result that the words are arranged such that they co-occur with similar ones is therefore not compatible with a linguistic system. An English text with similar features would mainly consist of words similar to the words "the", "and" and "to". Additionally, a word "the" would co-occur with words like "khe", "phe", "fhe", "tha", "tho", "thy", "thee" and "thee". The features described above point to a text generation algorithm. (Timm, 2016 7)

Linguists are able to mathematically measure the amount of information per word (bits) in a given language and assign it a numerical value. Vöynichese, when measured, received a
score situated between English and Chinese, putting it in the range of natural languages (Montemurro 2). Contrast this with yeast DNA and the Fortran programming language, upon which these measurements were also run and returned much lower scores (Montemurro 2-3). Another property that sets Voynichese apart is its statistical differences from randomly generated text, ritualistic incantations such as magic words, and pathological speech patterns (such as those associated with schizophrenia) which allow us to rule those possibilities out (Rugg 34).

Despite this unique collection of properties, Voynichese does still exhibit some linguistic regularities that would be expected of a natural human language. Certain letter combinations don’t appear (Rugg 34), and it follows Zipf’s law\(^{10}\) (Timm, 2016). Because it follows Zipf’s law, “the probability that the Voynich text resulted from some kind of stochastic process is drastically reduced” (Montemurro 2).

**Other Codes**

**The Decipherment of Linear B**

Linear B was a script discovered on tablets and pottery found in Greece dating back to the Mycenaean period, which dated from approximately the 16th to the 12th centuries BCE (Chadwick 1987, 6). Beginning around 1900, archaeologists began unearthing ancient tablets on the Greek island of Crete, particularly around Knossos, onto which two previously unknown scripts were carved. The first, called Linear A, was comprised of simpler shapes and figures, and there were relatively few examples of it. There were many more surviving examples of the more complex second script, deemed Linear B, which was dated to around 1450-1375 BCE (8).
Both Linear A and Linear B were totally unknown; they certainly did not appear to be any known form of Greek, and they predated even the earliest Greek alphabet by centuries. Interestingly, examples of Linear B were found exclusively on or near the ruins of ancient Mycenaean palaces, in contrast with Greek, which could be found anywhere in Greece (10-11). Archaeologists and linguists alike were baffled; while the location and dating of the surviving Linear B artifacts suggest it could be the language of the pre-Greek Mycenaean civilization, this hypothesis would have been impossible to confirm, and practically nothing was known about whatever language the Mycenaeans spoke. Scholars could not even agree that Linear B represented a unique language at all, as there existed the possibility that it could just be an alternate form of another known language. All of the confusion and uncertainty surrounding the origin of the Linear B script led to the language itself going undeciphered until 1952, when linguists Michael Ventris and John Chadwick finally solved the mystery.

But how does one take a totally unknown language and completely decipher it? The process is rife with difficulties, the first of which being the simple fact that no text like Linear B had ever been discovered before. There was no prior work done, and no foundation to work off from - any and all work done would be completely from scratch. It was not even known what language Linear B was supposed to be; the best guess was some previously unknown Cretan script, but nothing more than guesses could be established, especially since there were so few existing written records of that time period to provide any sort of contextual clues. And while, as Chadwick explains in his book, "the first step in trying to decipher an unknown script is an analysis of the texts" (12), the Linear B corpus was small, which makes any kind of statistical analysis more difficult on account of a smaller data set. In addition, it is important to remember
that all Linear B text was in the form of physical carvings, which layered on the additional
difficulty of interpreting the scribe's handwriting. Two symbols which look alike could be the
same symbol, or they could be completely different, but since there is no way to know that in
advance, both possibilities must be taken into account. There is also the possibility of symbols
with capital or final forms.

First and foremost, archaeological and linguistic methods were both crucial in the
decipherment of Linear B. Each can only go so far on their own, but together they can paint a
much clearer picture. Once a decent number of tablets had been discovered, the linguistic work
could begin. Out of all the symbols, numerals were the first to be identified. They were easiest
because they were on “virtually all tablets...and there are examples of addition to verify the
system” (13). The "examples of addition" were verifiable because once the numerals were
discovered, the next step was to look for symbols that only ever appeared with numbers. Sure
enough, symbols of such a type existed, and these soon became clear as mathematical operations
(e.g. addition) and units of measurement (13-15). On many tablets, the numerals and associated
symbols often appeared next to a particular kind of ideogram called a "pictogram", which is a
straightforward ideogram that looks like what it represents (For instance, 𒈴𒌷 meant "horse" and
𒅤 represented a particular kind of jar). The Linear B tablets were therefore deduced to be “lists
of men and women, livestock, agricultural produce and manufactured objects. Without any
knowledge of the language, we can still give a useful account of the subject matter of the
records” (15). And with this breakthrough, the nature and purpose of the tablets became known,
and some parts were able to be 'translated,' or at the very least understood. However, most of the
symbols on the Linear B tablets were not ideograms or numerals, and so remained unknown.
From this point on a more rigorous linguistic approach became necessary. Despite the relatively limited size of the corpus, some statistical analysis was still possible. Ventris and Chadwick performed a frequency analysis and made note of not just how often each symbol appeared, but also how commonly each symbol is the beginning, middle, or final symbol in a word. Organizing the data in this manner allows room for existing patterns to emerge, and

… at this stage it is useful to compile an index of sign-groups, not merely to find repetitions of groups, but also to discover groups which begin alike but have different endings. Likewise a reverse index, that is to say, one in which the groups are arranged in the order of their signs working from the end of the word, is useful to find groups that have the same ending” (17).

That is, this technique can be used to identify potential word roots and word endings, and "if the language of the inscription had been known it should not have proved too difficult to find values which would give appropriate words" (17). However no such assumption could be made because Linear B was still an unknown language. It is at this point that a decipherment effort could potentially reach a dead end. There is no more linguistic data to collect, and there are only so many known existing tablets. The process from here on ventured from the foundations of linguistic data to the realm of linguistic hypothesis. Specifically, the working assumption that was made was that Linear B could be an inflected language (a language with words composed of word stems and endings which change based on context), and so the properties of inflected languages would also apply to it. They took words that had identical roots but different endings that were used in similar situations to potentially be different genders of the same word, differentiated by the vowel sound at the end. Many of pairs of this kind existed, and using these pairs they formed a grid which associated each symbol with a particular consonant sound and a particular vowel sound. This grid, in other words, was a potential syllabary of Linear B. This
syllabary was then compared against an early syllabic Greek script. Similar 'words' for similar concepts emerged, including names of specific places, but this was not a universal phenomenon between the two scripts. There was not enough evidence to support this as much more than a coincidence.

Then another tablet was discovered, a tablet that essentially confirmed this hypothesis. This tablet was called the "Tripod Tablet," so named for the ideograms of tripod structures appearing on it (𐃠); reading the Linear B off the tablet and comparing it to the syllabary produced an almost perfect match to the syllabic Greek word for "tripod." Other specific structures appeared in the text as ideograms, and fortunately those were also accompanied by Linear B words which correlated exactly to the syllabic Greek. This was so specific that the probability of it being a coincidence was extremely low. Thus, Linear B was finally revealed to have been an early Mycenaean syllabic script of the Greek language (18-21).

The story of the decipherment of Linear B is a fantastic example of successfully deciphering a completely unknown language, and there are certain fundamental principles that can be learned from it. First of all, it demonstrates that the size of corpus is immensely important. More data allows for a closer analysis, and any part of a corpus might contain something key to the decipherment. If, for instance, the Tripod Tablet had never been discovered, then the Linear B conclusion would still only be conjecture. Next, it shows that an unknown script could actually be a previously unknown form of a known language. While the archaeology may hint at it, it is not at all obvious that Linear B is a syllabic-mixed early script of Greek. However once that connection was made, the rest fell into place. The decipherment effort also shows that there are multiple elements of a text to analyze, including not just the symbols, but also the behavior of the
symbols, the appearance of the symbols, and the context of the symbols. There are many possible approaches to a problem such as this, and there is no possible way to know in advance which approach (if any) will succeed. It may be necessary to create a hypothesis and proceed on the faith that it will succeed. If a hypothesis does not reach a dead end, then a series of multiple hypotheses may follow from each other, situated overall on a foundation of the data. With luck, the hypothesis will consistently hold.

**The Partial Decryption of the Kryptos Monument**

Kryptos is an art installation that the CIA had commissioned to be installed at their headquarters. It was created in 1990 by artists James Sanborn and Ed Scheidt. The sculpture itself resembles a large scroll made of copper, displaying sets of letters. On one side is a series of coded messages, and on the other side is a decryption key.

*Kryptos*
It is not obvious upon first glance what the encoded message is, or how one would decipher it. As of 2017, only three out of the four sections have been decrypted. One of the first people to publicly decode the sculpture was an amateur cryptanalyst named David Stein. Stein released his solution in 1998, 8 years after the monument was erected. The following information is from his account solving Kryptos.

The first difficulty in deciphering this code was beginning, since there were no hints. A first step would have to give information that is useful regardless of what encryption method was used. Stein decided to start by performing a frequency analysis of the Kryptos text, which is
"counting up the number of times that each letter of the alphabet occurs throughout the message" (Stein). Because it is possible that "different sections of the Kryptos code could...have been enciphered with different coding schemes...it would be necessary to perform frequency counts on each part separately” (Stein). By seeing how each character’s frequency evolved over the rows, any sudden changes would signify a potential change in pattern, and therefore help divide the code up into smaller parts.

*Results of Stein’s Frequency Test for the Letter “J”*

Stein performed a frequency analysis on all 26 letters for each line in the sculpture. Here is his result for the letter ‘j’. As the diagram shows, the frequency drops off at line 15 before it picks up again at line 27. Because ‘j’ is one of the less frequently used letters in English, this
suggested to Stein that the code was in fact divided into different parts, and that these parts were encoded by different methods.

Upon closer inspection, Stein noticed question marks in the text which seemed to divide the two halves of the text into subsections. Stein chose the second of these subsections to start with, because he noticed that the substring “DQM” repeats twice within its first 11 letters. Comparing the frequency of letters in English with the frequency of letters in section 2, Stein noticed that the letter frequency for this section of the code, was flatter and more evenly distributed. According to Stein,

“This is exactly what [one] would expect to see for a polyalphabetic substitution code, where the distributions for several different alphabets would tend to average out when combined. Because the modified Vigenere Tableau – a polyalphabetic substitution code – is inscribed on the Kryptos sculpture, it made sense to try this first as a working hypothesis for this part of the code.”

Working under the hypothesis that this section was encoded with a Vigenere cipher, Stein’s next task was to determine the length of the cipher’s keyword (Stein determined that it is unnecessary to find the actual keyword; knowing the length is sufficient to find a solution).

It is possible to find the length of the keyword using a cryptanalytic tool known as a text’s Index of Coincidence (I.C.). The Index of Coincidence measures how likely it is that two letters from two different texts (in this case the different alphabets) occupy the same position. The closer this value is to 0.066, the more likely the cipher is to be a mono-alphabetic cipher.

Stein tested the IC of the section for multiple different potential keyword lengths and the results suggested a keyword 8 letters long. Ordinarily, the process from this point would be straightforward; perform a frequency analysis on each of the 8 alphabets, match the highest occurring letters to ‘e’ (the most common English letter), and infer the rest. However, this section
of text was too short to perform that type of analysis on and get any meaningful results. With only 15 letters per alphabet, there was insufficient data for a meaningful answer. A more exploratory approach had to be taken.

Stein went column by column, printing out all 26-different alphabet-shift possibilities and finding the relative overall frequency of the English letters of each. By process of elimination, one of the columns must consist of every 8th letter of the plaintext message. He found one starting with ‘T’, and since “THE” is the most common English word, he tried out setting the 2nd alphabet to give ‘H’ and the 3rd to give ‘E’. This approach seemed to produce results, and the rest of the puzzle fell into place. Using similar techniques, Stein found that the first section of the puzzle seemed to be encrypted using the same modified Vigenere technique with a different keyword.

Stein found frequency analysis to be a good starting point, so he applied it again on the third section of Kryptos. He found a letter frequency similar to English, which made a modified Vigenere cipher unlikely. This result suggests that a transposition cipher was almost certainly used. Once Stein realized that this was a transposition cipher, he was quickly able to find a solution.

The fourth section of Kryptos is still unsolved. There were some misspellings in the translations, including “Undergrund”, “Desparately”, and “Iqlusion”. It is unknown what the function of these typos are, if there even is an intended one (The artist has hinted that these might be relevant to decoding the fourth section). The sculptor admitted in 2006 that Stein’s decryption actually had an unintended typo due to a mistake he had made in the sculpture.
There are important principles of decryption that we can extract from David Stein’s Kryptos story: like with Linear B, we once again see the utility of statistical analysis, and we once again run into the problem of a small corpus size preventing analysis beyond a certain point. We see that different parts of a message can be encoded in different ways, and if a mistake was made in the encoding process, the information encoded in the message could be irreversibly altered or lost entirely. We also see again that sometimes the only way to make progress is to make a hypothesis based on the data and proceed on the assumption it is correct.

Investigators of the Voynich Manuscript

Tables and Grilles: The Makings of a Hoax?

Ever since the Voynich Manuscript's rediscovery in 1912, it has been theorized to be a hoax. These theories had trouble gaining traction because there seemed to be more evidence suggesting its authenticity. Earlier in its life, the manuscript had allegedly been in the possession of the Elizabethan scholars and known hoaxers John Dee and Edward Kelley (Rugg 33). Dee and Kelley had previously forged writings written in the strange 'Enochian' language, which they claimed was literally the ancient celestial language spoken by angels. One hoax theory claims that they similarly forged the Voynich Manuscript in order to sell it to Emperor Rudolph II (35-36). There are problems with this theory. For one, Enochian was easily debunked and shown to be linguistically inconsistent with many of their claims. It was too simple. The complexities of Voynichese on the other hand make it completely different from Enochian. It has some linguistic properties like other natural languages and some completely unique to it. If the Voynich Manuscript is a hoax, then it is an extremely sophisticated one; "it has generally been assumed
that a hoax containing these features would take an enormous amount of time to generate, and
would not be economically viable for a hoax perpetrated for financial gain”(34).

Under the assumption that the manuscript is not a hoax, it is natural to assume that it does
contain meaning, but that the meaning has been encrypted in some way. Given its age, it is
unlikely that Voynichese is a simple cipher; modern cryptographers have been able to decipher
every other known pre-modern cipher within days. In addition, some of its linguistic properties,
such as word co-occurrence are not accounted for (33). Other theories suggest that the Voynich
Manuscript at one point contained meaning, but has been rendered undecipherable because the
scribe either made a lot of mistakes or was using a one-way encryption technique. However
neither of these explanations account for its "linguistic regularities" (34).

In his article, Rugg describes a way that the Voynich Manuscript could feasibly have been
created as an asemic (without semantic meaning) hoax in a reasonable amount of time in a
manner that accounts for its unique linguistic features, using only techniques available during
Dee and Kelley's time.

The first step in devising this text generation method was in the observation that "the
number of rows [40] in Dee and Kelley's table [used to generate Enochian text] is near the
maximum number of lines of text on a page in the manuscript" (36). Although Enochian and
Voynichese are very different, it is possible that the text of the Voynich Manuscript may have
been generated by a similar technique (35-36).

Rugg's proposed method uses a combination of this kind of character table with a
modified Cardan grille. For simplicity's sake, the following references to characters may be more
easily thought of as syllables, but the principle holds in either case.
The first step is to make the table. Construct the table such that the columns are organized into sets containing 3 adjacent columns. Characters appearing in a cell under the first column represents a word's prefix, those under the second column an infix, and those under the third column a suffix. Each set corresponds to a word, so the number of sets in a table represents how many words can be generated at a time with a single use of the table. There will be multiple rows, such that each row could correspond to a line. Therefore a table could correspond to a page. Populate each cell such that some contain characters and others are blank.

The next step is to prepare a grille. Like a Cardan grille\textsuperscript{22}, this grille will have holes at specific locations on its surface through which a single character can be seen each. The holes will be adjacent in sets of 3, such that each set of holes will reveal a prefix cell, an infix cell, and a suffix cell of a single word. While each grille will have three holes adjacent to each other vertically, each grille will have the holes at different heights (i.e. rows) relative to each other. There can be any number of different tables and grilles.

To begin to generate words, overlay a grille onto a table so the prefix, infix, and suffix holes reveal a prefix, infix, and suffix of a single word's column set. Combine those characters in prefix-infix-suffix order to create a word. If the prefix hole resides over a blank cell, then the word will simply be infix-suffix. Likewise for blank infix and suffix cells.

Repeat this process, placing the grille in a different position each time, in order to generate more words. Moving the grille in a perfectly systematic way (such as by always moving it down 1 row) makes the resulting text too regular, so a more complex algorithm for moving the grille must be used in order to create variance. A large number of variances in tables, grilles, and
grille-moving algorithms means that a large variety of fake text could be systematically produced very quickly (36-38).

Upon using this method to replicate the process of generating fake Voynichese text, Rugg had this to say:

Once the table and grilles have been set up, which can be done in two or three hours, this method produces text as fast as it can be transcribed. The total time taken to illustrate a replica page, generate text and transcribe it was consistently between one and two hours for a normal page (38).

That is, this process only uses methods available during the time of the manuscript's creation, and it allows a work of such a size to be created within a reasonable timeframe. Interestingly enough, this method can also account for certain linguistic properties of the Voynich Manuscript. Rare or absent character combinations in the text, for instance, can be explained with this method of text generation. Depending on the specific setup of characters in the table, and the orientation of the holes in the grilles used, some characters may be oriented relative to each other in such a way that they never appear as part of the same word (38-39). Other constraints can be artificially imposed as well, such as color coding the characters and creating a rule for how certain colors may or may not be allowed to go together (39). This method also predicts that, “other things being equal, words produced from the same column of a grille will be more likely to resemble each other than words produced by different columns” (40), and that common words with common characters are more likely to co-occur. Likewise, rarer words and words with rarer characters are less likely to co-occur (42).
One finds that even the slightest changes leave measurable statistical effects on the text. The grille-moving algorithm used is an apparent one, and another is the "edge effects" that occur from characters being placed in particular places in the tables. Basically, the fewer instances of a character in a table, not only will it obviously be rarer in the text, but it will also be more restricted on where in a line it will occur. That is, the rarer a character is, the more specific limitations it will have on how and where it can appear (For instance, the ‘m’ character is almost always a character at the end of a line. This could indicate was placed in a particular cell in the generating table which would produce such an effect). Likewise, edge effects can also affect the frequency of certain characters occurring more often in the beginning or end of words and/or lines (41). Another type of edge effect relates to characters placed in the top and bottom rows of the tables. For any particular grille on any particular table, unless all 3 holes are on the same row, there will be characters which simply are not used (41).

There are other properties of the text that hold under this table and grille hypothesis. If a hoaxer wanted to make section titles or proper nouns stand out linguistically from the ordinary text (in order to fool anyone who might scrutinize their work, since, like with the Rosetta Stone, proper nouns are often “a potential way into the document”), they could have a special table reserved for just that (42-43). In a similar way, there could be a special table with a lot of empty cells when a shorter word is needed to fit at the end of a line. The table and grille method would also explain the significant differences found between Voynich A and Voynich B. Two scribes working in parallel with different tables and grilles would perfectly explain the different handwritings and statistical properties (43).
Rugg puts forth a strong argument for the Voynich Manuscript being a hoax that was developed using a simple text generation algorithm. It explains some of the unique linguistic features of the text in a way it realistically could have been done at the time of the manuscript's creation. If one assumes that the table-and-grille method (or something similar to it) was used, then it is also necessary to accept the Voynich Manuscript as meaningless asemic writing because while grilles can encode meaningful text (their original application as a steganography tool) there has been no evidence found to support a table-and-grille method of plaintext encryption (43).

Rugg's argument is plausible, and he has demonstrated that it is possible to generate meaningless text with interesting statistical properties. There is a chance that the Voynich Manuscript is a hoax, but this claim is ultimately unverifiable with the physical evidence that currently exists, even if the results can be recreated.

An Amateur Attempt

In September 2017 The Times Literary Supplement ran an article written by Nicholas Gibbs where he claims to have solved the mystery of the Voynich Manuscript. He claims that each letter of Voynichese is actually a ligature of Latin script, and the “words” in the manuscript are actually abbreviations. Taken altogether, he claims the manuscript is a list of recipes. Gibbs claims that the manuscript is “a reference book of selected remedies lifted from the standard treatises of the medieval period, an instruction manual for the health and wellbeing of the more well to do women in society, which was quite possibly tailored to a single individual” (Gibbs). This is not an implausible idea per se, but Gibbs’s conclusions were not novel, and they were poorly fleshed out. While there is no way to confirm Gibbs’s idea, this was a plausible conclusion that many researchers already believed.
One of problem of many with Gibbs’s translation was that the Latin his technique produces does not make much sense. It is grammatically incorrect. This raises the question: why would someone encode information that does not make sense in its original language? Another problem is that his argument rests on a weak foundation. He claims that the text contains no actual words for any illness, medicine, or plant. All references to these are instead references to the manuscript’s index (for the sake of brevity in what is already claimed to be an abbreviated system). This alleged index conveniently happens to consist of the exact pages that are missing from the manuscript.

As far as results go, Gibbs released only this single, small, grainy image of the results of his technique as applied on only two lines of the manuscript.
This is just one example of many, but many amateur “solutions” fall into the same traps. Those traps are an inherent difficulty to verify the solution, and an overabundance of jumping to conclusions.23

**Conclusion**

Looking at the successes of Linear B and Kryptos, we can infer that increasing the probability of a successful decryption or translation requires these three things:

1. A large enough corpus. Linear B went undeciphered until more text was discovered. The fourth section of Kryptos is still unsolved, due partly to the ineffectiveness of statistical analysis on account of its short length. Hypothetically, if more text written in the Voynich script were to be discovered, then a solution may be found. However until then we can only derive so much.

2. An external reference point. This is to check for correctness. Linear B had syllabic Greek and Kryptos had the artist. The Voynich Manuscript is the only text of its kind, and the author, whoever they may have been, is long-since dead. Even if a correct solution were to be found, how would anyone actually know that it is correct?

3. Luck with a data-grounded hypothesis. This is the methodology that worked for Linear B and Kryptos. The fewer jumps in reasoning one has to make throughout the process of working through a hypothesis, the stronger the hypothesis is. In Kryptos and Linear B, only the initial steps had to be assumed, and the rest followed from that.

Will the Voynich manuscript ever be decoded? My personal answer is: Not in its current state. The fact of the matter is, there is too small of a corpus, and no way to verify correctness. In
addition, the possibility exists that the scribes could have made a transcription error, and the message could have been accidentally lost. The strongest argument I have read appears to be the hoax hypothesis. Researchers have been able to demonstrate a method, possible during the 15th century, that could generate an amount of text with the Voynich Manuscript’s unique linguistic properties in a short enough time for it to be financially viable.
Works Cited


Rugg, Gordon. "An Elegant Hoax? A Possible Solution To The Voynich Manuscript."


End Notes

1: Kircher’s translation was completely incorrect. He was believed at the time, but hieroglyphics were not actually properly translated until the discovery of the Rosetta Stone.

2: Picture of the “Herbal” section. Includes some of the “composite” plants.

3: Picture of the “Astronomical” or “Cosmological” section. Includes what may be the sun.
4: Picture of the “Biological” section. Includes nude women bathing in an organ-shaped body of water.

5: Picture of the “Pharmacological” section. Includes what may be medicine jars.
6: Picture of the “Recipes” or “Stars” section. Note the stars drawn in the margin.

7: I say “somewhere between” because certain characters are difficult for researchers to differentiate due to the scribe’s handwriting.

8: Gallows letters. They much more commonly appear at the start of lines than in the middle or end of lines.

9: The Voynichese alphabet. Underneath each character is a common “reference letter” used when transliterating into the Latin alphabet.
10. According to Wikipedia, “Zipf's law states that given some corpus of natural language utterances, the frequency of any word is inversely proportional to its rank in the frequency table. Thus the most frequent word will occur approximately twice as often as the second most frequent word, three times as often as the third most frequent word, etc.: the rank-frequency distribution is an inverse relation.”

11: As far as written scripts go, there are three different kinds: phonetic, ideographic, and mixed (Chadwick 1987, 12). In a phonetic system, the symbols represent the "sounds of the language" that the script represents (12). That is to say, there is a direct relationship between the organization of the symbols in a word (i.e. its spelling) and the pronunciation of the word, and
there are standard conventions in place to account for any minor pronunciation differences among individuals (12). Technically speaking, anyone who can properly read a particular phonetic language can speak it, even if they do not understand what they are saying. Phonetic systems can be further broken down into two different types: syllabic and alphabetic. In a syllabic script, each symbol represents a syllable, and the different combinations of these syllables form words. In contrast "alphabetic scripts aim for the ideal of one sign for each sound" (12). Not every alphabetic language meets this ideal however, which necessitates the existence of diacritical marks, diphthongs, and digraphs.

The other major kind of writing system is ideographic scripts, which "have basically one sign for each word, and this sign usually represents the meaning of the word, not primarily its phonetic form" (12). In contrast to the syllable or the letter which both stand for a sound, the ideogram - the grapheme of an ideographic script - is a semantic unit, representing a single concept. Because of this it is possible to know how to read a language represented with an ideographic script without being able to speak it, and vice versa. This also means that there are significantly more symbols in an ideographic script than in a phonetic one. While this makes ideographic scripts much easier to recognize, it makes them much more difficult to learn to read and write (12-13). Ideograms also have an advantage in that they can be "easily transferred from one language to another," such as Japanese kanji borrowing from Chinese, because the phonetics are not connected (12-13). Perhaps the best example of this phenomenon is Arabic numerals (12-13). Take the symbol "2" for instance; regardless whether one calls this symbol "two" (English), "duo" (Latin), or "zwei" (German), the semantic meaning as the second positive integer remains the same.
In addition to phonetic and ideographic scripts, there is also what is known as a mixed
script, which as its name implies is a script which combines both phonetic and ideographic
components (13). Some of the better known examples include Japanese, the ancient Hittite
language, and ordinal constructions such as "1st", "2nd", etc. (13).

12. A larger view of Kryptos:
A Vigenere tableau is a 2-dimensional table where the x and y axes are both the English alphabet which is used to help decrypt a Vigenere cipher. This table is also sometimes known as a Tabula Recta. A Vigenere cipher is an extended form of the Caesar cipher, which is one of the earliest and simplest encryption schemes. To encode a message in this manner, the encoder chooses a number and shifts every letter of their message that many spaces forward in their alphabet.
A Vigenere cipher is a series of different Caesar ciphers applied to the same message based on a keyword. Here's how it works: The encoder first selects a word which the cipher will be based off of. This is the keyword. This keyword is repeated to a length equalling that of the message to be encoded. This repeated-keyword string is placed next to the message, and each letter of the message is shifted (as in a Caesar cipher) by an amount determined by the corresponding letter of the repeated-keyword string. The plaintext letter is shifted such that an 'A' would be shifted to become the corresponding letter of the keyword, a 'B' would be the letter immediately following the corresponding letter of the keyword in the alphabet, and so on. In other words, a Vigenere cipher can be seen as a cycle of different Caesar cipher shifts on a message. Assuming a regular Vigenere alphabet table, an alternative way to think about it would be to assign each letter of the alphabet a letter value, with 'A' equalling 0, all the way up to 'Z' equalling 25, with the Caesar shift being equal to each corresponding number.

Here is an example:

Encrypting the word “ENCRYPTION” with the keyword “VOYNICH”:

“ENCRYPTION” is a longer word than “VOYNICH” so the keyword is going to have to repeat.

If we assign each letter of the repeated-keyword string a number value based on its order of the alphabet, "VOYNICHVOY" would be (21) (because of V’s place in the alphabet) (14) (because of O’s place in the alphabet) (24)(Y) etc.
Thus, each letter of "ENCRYPTION" would be shifted by a different amount. 'E' is shifted 21 places to become 'Z', 'N' is shifted 14 places to become 'B', 'C' is shifted 24 places to become 'A', 'R' is shifted 13 places to become 'E', 'Y' is shifted 8 places to become 'G', 'P' is shifted 2 places to become 'R', 'T' is shifted 7 places to become 'A', 'I' is shifted 21 places to become 'D', 'O' is shifted 14 places to become 'C', and 'N' is shifted 24 places to become 'L'.

Thus, the final encoded message would read ZBAEGRADCL.

It becomes clearer from larger examples that the letter frequency of messages encoded by Vigenere cipher is different than that of ordinary English. Stein's letter frequency analysis results showed differences from ordinary English, which supports the idea that the first part of Kryptos may have indeed been encrypted in this or a similar manner. As stated previously, Stein had noticed a "modified" Vigenere tableau on the monument. The specific modification is the insertion of the word "KRYPTOS" in each line of the otherwise regular Tabula Recta.
15: Example Index of Coincidence calculations

(https://www.elonka.com/kryptos/mirrors/daw/boxb.jpg)

16: It turns out that the first word was actually “They”, not “The”. Stein just got lucky.
17: Plaintext message of the final message of the top section.

![Figure 18. Completed plaintext for top section of Kryptos code.](http://www.elonka.com/kryptos/mirrors/daw/figure18.jpg)

18: Notice some words are misspelled, such as “Undergruund” and “Iqlusion”. This is possibly intentional.

19: A transposition cipher is a cipher by which a message is encoded by scrambling the positions of each character in the plaintext. One common way to encode something using a transposition cipher is to write the message out in columns, rearrange the columns, then write it out by rows, as seen in this picture:
### Transposition Cypher (Example)

**Plain:**

```
FOURSCOREANDSEVENYEARSAGO
```

```
1 2 3 4 5 3 2 4 5 1
FCNER NCERF
OODNS DONSO
URSYA SRYAU
REEEG EEEGR
SAVAO VAAOS
```

**Cypher:**

```
NCERFDONSOSSRYAUEEEGRVAAOS
```

*Figure 21. Example of a basic transposition code.*
20: Stein’s solution was based on figuring out the likely dimensions of the rows and columns.

The final message, a quote from the account of the discovery of King Tut’s tomb, is as follows:

```
SLOWLY DESPARATLY SLOWLY THE REMAINS OF
PASSAGE DEBRIS THAT ENCUMBERED THE LOWER
PART OF THE DOORWAY WAS REMOVED WITH
TREMBLING HANDS I MADE A TINY BREACH IN THE
UPPER LEFT HAND CORNER AND THEN WIDENING THE
HOLE A LITTLE I INSERTED THE CANDLE AND
PEERED IN THE HOT AIR ESCAPING FROM THE
CHAMBER CAUSED THE FLAME TO FLICKER BUT
PRESENTLY DETAILS OF THE ROOM WITHIN EMERGED
FROM THE MIST. CAN YOU SEE ANYTHING?
```

Figure 25. Correctly ordered text for Part IV.
21: Notice also the final words in this part of the message “ID by rows”. This is not the intended plaintext, and was produced as a result of a typo made in the sculpture’s construction. Section II originally had an ‘X’ at the end of one of the lines, (signifying punctuation), but Sanborn deleted it “for aesthetic reasons” in order to keep the sculpture “visually balanced”. This had the unintended side effect of making the final 8 letters decode to “ID BY ROWS” when they actually are supposed to be “(X) LAYER TWO”

22: A grille is a sheet with holes cut out of it, like in the following image. A Cardan grille is a particular kind of grille used to encode messages. The assumption made is that the recipient knows there is an encoded message and has a matching grille.

![Cardan Grille Image](https://en.wikipedia.org/wiki/Cardan_grille#/media/File:CardanGrille.png)
23: In addition to all the problems, only a couple paragraphs of the article actually detail this attempt, while most of it is an autobiographical piece that feels like an attempt to sell a book. That’s just my opinion, and while that does not weaken his argument *per se*, it certainly doesn’t add credence to his academic thoroughness.