Benjamin Tepolt
MATH-341 : Cryptology
March 23, 2013

Project Report

Voynich Manuscript Cryptanalysis Software

Introduction

The Voynich Manuscript is certainly one of the world's most mysterious documents. It is written in a language nobody understands, filled with the drawings of fantastic-looking plants, and a slim section of twenty pages in roughly the center of the book features naked, dancing nymphs. Carbon-dating places it to the early 1400's, hand-writing analysis suggests at least six authors, and there are numerous letters between investigators researching it during the 1600's. Is it a hoax, or is it really just something which is not yet understood?

The story of the document's discovery is almost as interesting as the document. Michał Wojnicz (Voynich) escapes Siberia after being sentenced to penal servitude for his involvement with Socialist revolutionaries in Poland. Ten years after this, he marries another revolutionary, Ethel Boole -- the daughter of George Boole. Like much of mathematics, it's nice to be in such good company. The manuscript bearing Voynich's name is discovered in 1912, along with a collection of 30 other manuscripts that weren't fascinating enough to make history. Those who were so willing to see a new world through revolutions were able to look at the Voynich Manuscript and believe that it has a real meaning.

Cryptologists from World War 2 were unable to crack the Voynich Manuscript, but computers were much more limited back then. The software for the cryptanalysis system will be designed using PHP.

Working Theory

It is premature to make a working theory for a text before doing any cryptanalysis whatsoever. However, before I did any programming, whether it was investigation or decryption, I looked through all of the drawings in the book from the first to last page. I sought a logical explanation of the drawings alone.

Like most cryptological techniques, I'll begin by describing the most obvious drawings first before moving on to the less obvious, hoping that the former can aid in the latter. The first section of the book is an herbal section, containing plants along with text that most likely describes them, in an

encyclopedic manner. I'm convinced that the text describes the plants, because each drawing of a plant is accompanied by maybe two paragraphs of text and tremendous amounts of white space -- whereas in the last 10 or 15 pages of the book, the author is making as use of white space as economically as possible.

The first herbal section is followed by an astrological section, then the nymphs dancing in pools of green and blue liquid, another herbal section of the same style as the first, concluding with about 15 pages of pages without any drawings at all. The astrological section, which contains the signs of the Zodiac representing parts of the year, seems like the next most-obvious part after the herbal section. It is likely used as a calendar to determine when when a month has arrived or past, using fixed stars over the horizon to determine the earth's tilt.

Some of the herbal pages contain one or two pages that are full of diagrams of baby plants or plant roots, suggesting that it wasn't just a guide for finding plants in the wild, but for transplanting them to a private garden for controlled production. During the 1400's, there were many places throughout Europe that needed constant gardens, whether a monastery, a castle, or a church, so it's nice that historically the book makes sense. The astrological sections probably indicate when certain plants are available in the wild at their various stages of growth.

Then there is the most interesting and least explicable of the drawings found in the book: the dancing, naked nymphs, dancing inside of tubes or pools of liquid (theoretically), then followed by an infant plant section, with each collection of infant plants drawn next to what looks like a Chinese incense burner, which would've been in production from a thousand years before the book to just about the time the book was written. The curled, triple-legs of these incense-burners is strong, but not definitive, evidence of their origin.

The naked nymphs are not literally naked nymphs -- they are drugs and the incense-burners are stills. This is why they are dancing in one tube that is full of a green liquid (the reagent) that finally leads to another tube full of a blue liquid (the distillate). Alcohol: the word is Arabic, with "al" meaning "the" and "cohol" meaning "spirit," truly indicating the mindset of alchemists working with drugs. They were separating out spirits (as in bottled spirits; liquor) from plant material. One of the plant drawings resembles Milkweed, which apparently has some antihistimine properties. This explanation leaves a question, though: why are there drawings of about 20 different stills, when it's likely that alchemists wouldn't have had the skill to make 20 stills that functioned in different ways?

Preparation

Several minor things needed to be coded up first before handling any

investigation or decryption.

1) View Data

This page allows a user to view any page of the Voynich Manuscript, with both an image-file of the page and a text-format of the page, using the European Voynich Alphabet (EVA) format. EVA is simple an English-Alphabet to Voynich-Alphabet translation. Surprisingly, many of the characters match up quite close in appearance, like A, C, D, E, I, K, L, O, P, Q, S, and Y. I owe tremendous thanks to René Zandbergen and Gabriel Landini for coding up the 231 pages of the Voynich Manuscript, which easily would've taken me the entire semester to do.

2) Prepare EVA Data for Searching

The EVA data online is available thanks to Jorge Stolfi, Stanford University. However, each EVA page was its own page, and this required downloading 231 pages, which I did using a program not listed in this report. But then, each EVA page had editor's comments, which then needed to be stripped out. That is what the "Prepare EVA Data for Searching" did, since I wanted to examine word frequency and the like. Two versions of the data were created: one with just the Voynich words, each separated by spaces, and another that's just the same, except with a double carriage-return separating each paragraph. Both versions proved incredibly useful in all future work on this project.

3) Prepare Language Samples for Searching

Language samples seemed to be useful in helping me to analyze the Voynich Manuscript, since there's no logical reason for believing that the text was originally English and then encrypted. Language samples would be collected for 130 books in 40 languages, with the books coming from Project Gutenberg (for European languages) and translations of the Quran (the only way to obtain a long sample from every Middle-Eastern language). Samples were selected with an eye toward getting documents that were close to the length of the Voynich Manuscript, truncating the length when necessary.

There was also a difficulty with Chinese and Japanese: these languages do not traditionally have spaces separating the words. And there's no way to simply separate each Chinese glyph by a space, since a word can be one or many glyphs. For Chinese, I had to master the command line-based program often simply called *the Stanford Segmenter*, which uses two well-respected, Chinese dictionaries. And for Japanese, I had to master a similarly bare-bones program called *JVnSegmenter.* Like most open source software that's incredibly rare, there's little documentation and the only method of figuring them out (which I did) was the time-consuming

process of trial-and-error. This part fascinated me: few other cryptologists could've been able to do the type of analysis I'm doing, because these programs wouldn't have been available nor would Project Gutenberg's translations.

For the function, preparation here produced two versions: one version with all words separated only by a single space each, and another version that retained newlines/carriage returns ('\n' and '\r').

Investigation

The Voynich Manuscript is too orderly, too full of effort, too specific and exacting in its diagrams to be a hoax. But, on the other hand, I could simply have an overactive imagination. It's worth writing a few functions to determine whether the Voynich Manuscript is a hoax or not. I justified this reasoning by believing that even if the tests were inconclusive, they may reveal a pattern or some type of weakness that would allow me to try a decryption attack.

4) Unique Word Comparison of Voynich to Language Samples

For any document, you have words that appear only once and those that appear multiple times; I will term the former as "single-appearance" and the latter as "multi-appearance" words. You can have two percentage values: one is the total number single-appearance words divided by the total number of words, and the other is the number of single-appearance words divided by the total number of unique words. I have no idea whether a pattern for these values should appear across multiple samples of a single language, or even if it should. But I was quite amazed by the results.

After testing 130 language samples, I can say this: take any book from any language in the world, and the total number of single-appearance words will very likely be 60% to 80% of the total unique words. Bengali, Chinese, Italian, Old English, Middle English, and even Esperanto all exhibit this property. The Voynich Manuscript ranks at 69%, and the most distant distant languages from the Voynich Manuscript are Vietnamese at 41%, Japanese at 49%, Urdu at 57%, and Arapaho (Native American) at 81%.

It was pleasing to note that both Vietnamese (41%) and Japanese (49%) were closely related in their statistics. Similarly, Slovenian (71%), Bosnian (70%), and Turkish (69%) are close together; Spanish (72%) and Portuguese (70%) are very close to each other; and Inukitut (the Inuit language) (79%) and Arapaho (the Native American language) (81%) are also very close to each other. This encourages the hypothesis that languages are merely averages of other languages geographically nearby. With this fact and the analysis, we may be able to pinpoint

the geographic area where the Voynich Manuscript was written. Then we would be able to make a much more intelligent decision about the language of the original plaintext.

However, there was a bit of minor disappointment with all languages being grouped around 60% to 80%. It's far too specific. If the difference between languages was huge, then it would be very easy to use this statistic to definitively find out the language of the Voynich Manuscript. But the difference of the closest languages to the Voynich Manuscript is measured in tenths and hundreds of a percentage point. It's enough to determine that the Voynich Manuscript is very strongly a genuine document; but it's not enough to determine much more than that. We do know some languages that it is most likely not written in; but what language it *is* written in still eludes us.

Fortunately, there is the other statistic: the total number of single-appearance words of the total text, instead of divided by the number of unique words. This statistic produced significantly greater distances between the languages: all languages range from 5% to 40%. Since the median is much lower, differences are more meaningful. A 1% difference between 14% and 15% is more significant than a 1% difference between 80% and 81%. Languages formed patterns again by geography: Arabic (29%) and Hebrew (30%); Japanese (4%) and Vietnamese (3%); Serbian (24%), Czech (24%), and Turkish (23%); Tagalog (Philippine Language) (20%) and Cebuano (Philippine Language) (23%); Bulgarian (23%) and Romanian (23%); Friulian (Rare Italian Language) (18%) and Italian (20%).

I could map the globe with these values in place of the country names, with the vague imposition that there are no nations -- there are only slight variations in peoples' culture, measurable in percentages, from any square foot on this planet to any other square foot on this planet. Every culture is only a few degrees away from any other, and there are no real "types" of cultures. It's nice to find at least one conclusion that honored the interests Voynich probably displayed in this document.

The Voynich Manuscript's single-appearance words over total-words is 15% and over unique-words it is 69%. The closest matches for both of these values is Danish (14%, 67%); Swedish (17%; 70%); and Bosnian (17%; 70%). Denmark and Sweden are both in Northern Europe, but Bosnia is in South-Eastern Europe. This is definitely an anomaly that needs explanation. This gave me the idea for the next function to be built for the investigation.

5) Calculate Standard-Deviation and Mean for Language Samples

The statistic of measuring single-appearance words out of the whole text and out of total unique words isn't very much data. Why not create a frequency list for all words and then make this a normal distribution, using standard deviation and mean as the comparison value between different distribution curves? This should provide a much more meaningfully comparison of the Voynich Manuscript with the language samples. At the least, it may corroborate or contradict the conclusions reached in section 4.

The first comparison we can make is the standard deviation, with the single-appearance words used as our frequency points. Once again, languages show similarities to each other, such as Vietnamese (80.00) and Japanese (84.94); Arapaho (17.82) and Inukitut (8.34); German (23.31) and Hungarian (20.22); Polish (17.55) and Russian (18.29). The Voynich Manuscript has a standard deviation of 22.64, and its top five comparisons are Napoletano Calabrese/Naples Italian Dialect (23.11), Romanian (21.45), German (23.31), Greek (21.07), and Italian (25.68). All of the languages suggested by the previous test suddenly drop in relationship to the text comparison. In fact, Arapaho (Native American) and Icelandic show greater similarity to the Voynich Manuscript than Danish, Swedish, or Bosnian.

The situation is quite reversed when we make the comparison to the mean. Languages in this case again show tremendous similarity to the languages of geographically-close cultures. The lowest mean was Inukitut at 1.88% and the highest was Vietnamese at 14.01%, but the vast majority were all between 2 and 5%. The Voynich Manuscript has a mean of 4.5%. High-ranking languages that compare are: Breton/Celtic (4.41%), Danish (4.68%), Bosnian (4.10%), Chinese (4.41%), and Catalan/Barcelona Dialect (3.90%). It is unusual that Chinese ranks so high in mean, when it ranks so low in the comparison of the standard deviation (Chinese's 74.79 to Voynich's 22.64). Swedish also ranks relatively high at 86.94%, just like the results from section 4.

There is a lot of confusion between these results: the unique-word comparisons (section 4), the standard deviation results, and the mean results. However, there are several languages that can be isolated as probable languages: German, Swedish, Danish, and Norwegian, in order of probability. Only a cursory investigation will reveal that Swedish, Danish, and Norwegian are almost all identical in writing, with most differences in pronunciation. While Danish and Swedish ranks low on standard deviation and very high on mean, German ranks moderately high for all indicators and very high for one of them.

6) Reprogram Unique Word Comparison for Phrases

There were several modifications I had thought of to get more data out of

the Voynich manuscript without understanding any of the words. One was to make comparisons of phrases instead of words when doing the unique-word comparison, standard deviation, and mean calculations.

When applied to unique-word comparisons: 2-word phrases of Voynich best match Galician, Romanian, Arabic, Arapaho, and Inukitut; 3-word phrases match Frisian/Italian Dialect, Telugu, Finnish, Arapaho, and Hebrew; 4-word phrases match Telugu, Frisian, Arapaho, Hebrew, and Old English; 5-word phrases match Czech, Old English, Gaelic, Arapaho, and Hebrew. These results get more and more corrupted as the phrase length increases, because many of the words are actually words from diagrams that name stars or plants. These comparisons would end up grouping such words together, which would be unnatural in any book not written in this style. Plus, they are probably not as valuable in terms of results as the standard deviation, which covers the data in more depth.

The word-variable phrase statistics for Standard Deviation and Mean was never programmed, along with a few other modifications I had in mind: word and phrase statistics for chapter by chapter of the Voynich book (separating chapters by logical division of the drawings) compared to language samples and removing all words and sentences appearing in diagrams so that the Voynich Manuscript appears more like a book. I had spent a lot of time trying to determine whether the Voynich Manuscript was authentic or not, and during that, I gained some knowledge about what language its authors spoke, or what may be just as good -- where the book was written.

7) Investigation Conclusions

Our data suggests the languages of German, Danish, Swedish, Finnish, and occasionally Romanian and Breton (Celtic). We may not know who wrote the Voynich Manuscript, but we can give a good estimate of where it was written and the native language of its author. This is essential for any type of realistic decryption. If we didn't know what language the Voynich Manuscript was written in, how could we determine the accuracy of an attempted decryption?

Unfortunately, the only conclusive result is that the Voynich Manuscript is authentic. We know from handwriting analysis that there six authors, or possibly six scribes, who wrote the Voynich Manuscript. The only way it could be a fraud is if its creators were some unknown geniuses at mathematics and statistics, that they predicted functions like standard deviation and mean, and then created a plan where they each would produce some of the 40,000 words in a such a way as to adeptly satisfy the limits of statisticians' equations. An advanced understanding of linguistics, botany, and painting would also be required. I'll choose the theory that it's an authentic code, or a natural language, with some real meaning behind it.

Gordon Rugg in 2003 stated he had figured out that the Voynich Manuscript was a hoax based on his own "Verifier" approach, concluding that the book was gibberish. Andreas Schinner in 2007 corroborated this by calculating statistics comparing the Voynich Manuscript to a hoaxed document, which he claimed showed a great similarity. I still prefer my results calculated from 40 languages with 130 samples; this is part of the reason why I went in depth into studying the nature of language and geography in this section.

Decryption

The investigation was enlightening, but it felt like I was reaching the point of limited returns. However, at least now I know that what I am trying to decode does have meaning.

8) Page Associations

This was the first idea that came to me. Create a multiple sets of pages, with each set called an "association". And then associate pages together based on the content of the imagery. About twenty page associations were created, such as pages with baby plants, plants with berries, calendar systems, medicine jars, the stars, and naked women. Then I "run the page association", and it produces a sorted listed of the words that appear on all of the associated pages. The word "Daiin" almost always appears, but it is the most-appearing word in the whole text, so I assume it will mean something like "the".

Unfortunately, there were no promising results. However, all of the pages with the sun on them contain the word "Or", and, even more interesting, this word is spelled out several times as "Orrrrrrr" when appearing in an astrological chart, so as to accentuate the same word. Translate "Sun" to German and you get "So"; to Danish and Norwegian "Søn"; and to Swedish as "Sön". The most-occurring word "the" is "Die" in German and "Den" in Swedish/Danish/Norwegian -- almost a conglomeration of the Voynich Manuscript's "Daiin". But there is something awkward about Daiin -- it always occurs at the end of sentences and paragraphs, compared with "the" in the languages we are working with (as well as most of European languages).

"Ar" is another high-occurring word in the Voynich Manuscript, and this may be associated with the Swedish word for "is", which is "Är." However, the accent over the A can completely change the meaning of this word in Swedish. It is close to the Norwegian/Danish "Er" word for "is". It was difficult to find any other Swedish, German, or Danish words that very closely matched up with the Voynich.

9) Word-Related Pages

This is the opposite of a page association: instead of grouping pages by related imagery, I look at pages that share X-many words in common, selected from a list of the total words. Then I can decide what if there's any shared meaning between the two pages that might enlighten on the meaning of the word. There is a great depth to the distribution of these words, as something that appears in one of the pages of a plant may appear again only once on one of the final pages that's full of paragraphs and no pictures.

After searching through about 100 different word-related pages, from two pages that share just one word to ten pages that share one word, there were no conclusive results. One really interesting result was a word that appeared on only two pages: once floating near a diagram of a baby plant and once on a page with a gigantic plant drawn on it, and the two looked similar. Knowing just one word really didn't help in decrypting the rest of the pages. This gave me a new idea for decryption, though.

10) The "Named Diagrams" Decryption Hypothesis Set

Stage 1: Coding Description

A decryption hypothesis set is merely a collection of Voynich words and what they decrypt to. The Voynich Manuscript Cryptanalysis code was programmed to have multiple sets instead of just one, so that users can try multiple different decryptions. The user has the option to "Run Decryption Hypothesis Set Comparison." This will produce a list of all pages where the words of the decryption hypothesis set occur, sorted with most-decrypted pages at the top and least-decrypted pages at the bottom. Then you can click a page and it will replace each of the Voynich words with the word it decrypts to in the Decryption Hypothesis Set.

Here is the human algorithm: Guess one word, re-run the comparison, a different page pops to the top in the most-decrypted list, and then either guess another word or remove the previous word because it forces other pages to make no sense. That sounds easy, but there's nothing to start with -- except for the diagrams. So, my Decryption Hypothesis Set is very easy: use the diagrams. At first, I gave them the broad categorization of nouns, since why would someone label a diagram unless the label described the diagram? But then it seems much more logical, after some thought, to assume that the diagram describes the word.

Stage 2: Creating a Set

There are hundreds of diagrams in the book, and I run into an anomaly -- and it was tremendous. On page 142 ("f72v3"), there is a diagram of a woman holding a star, and it reads: "daiin chepy". But the word "Daiin" always occurs at the ends of paragraphs throughout the entire book, from beginning to end. I kept staring at the writing, until I realized it. It's not a diagram with named components -- it's a full sentence, with words interspersed between the small drawings of naked women holding stars, most likely done for artistic purposes, although there may be a concealment intention, as well.

However, on page 126 ("f68r2"), which contains a diagram of stars, the words in general do not have a high rate of occurrence at all. "Daiin" and the other popular words do not occur. So, some of the diagrams are genuine diagrams, the others are fake diagrams. It was nice to discover this while working through the early astrological section, as the mythical section with of nymphs is full of fake diagrams, while the botanical sections of the latter book are full of genuine diagrams. By the end, it had become very easy to identify a diagram that was just a sentence from one that was actually full of named parts. At that time, though, I had written descriptions for exactly 778 of the diagrams in the book.

Stage 3: Running the Decryption Hypothesis Set Comparison

If you take the 778 words from the Named-Diagrams Decryption Hypothesis Set and you use them to decrypt the Voynich Manuscript, then the entire text is 15.4212% decrypted. Using the Decryption Hypothesis Set Comparison function, I was able to look up all of the pages of the book that had some part decrypted, with the most-decrypted at the top. However, all of the pages appearing as "most-decrypted" were those pages that contained the diagrams. Many of the non-diagram pages had so little decrypted that it would have been impossible to make any sense out of the unencrypted portion.

The quest, then, was to find unencrypted parts most surrounded by decrypted parts. There is no sign of punctuation marks in the Voynich Manuscript, so there's no way to tell where one sentence ends and begins. So the challenge is actually a bit more difficult than just finding mostly-decrypted sentences; we must find mostly-decrypted paragraphs. Several interesting results came up when searching through the book for good samples. First, almost all paragraphs start with the same few letters: P, F, T, and K. And, secondly, very rarely is the first word one of the entries from the decryption hypothesis set.

On page 128 ("f68v3"), there is a diagram with several sentences spiraling outward from the center. One of these sentences, although small, has a 50% decryption. It is...

"okeol cheody qolkeey ocheky chody otchdy oko okos"

And with its partial decryption, it is...

"[Naked Woman Holding a Star] [Naked Woman Holding a Star; Baby Plant] qolkeey ocheky chody [Star; Magical Nymph; Alchemist's Tube] oko [Star; Root-Structure]"

The ";" is used to indicate that the Voynich word was used multiple times in different diagrams to identify different things. The word "cheody" is used to identify a naked woman holding a star on page 146 ("f73v") and to identify a baby plant diagram on page 179 ("f89r2"). This is quite unusual, and it happens quite a bit. The word "otol" is used to identify a star, air coming out of a tube, a naked woman holding a tube, and a baby plant. The word "okeody" is used to identify a green-blue tube, a star, three different naked nymphs holding stars, and a baby plant. This pattern appears again and again: stars, nymphs, plant.

Stage 4: The Theory

It would possibly make sense to combine the magical characters and the stars in a way as to identify the planetary objects as mythical beings. But once the plants are thrown in to the combination, nothing immediately comes to mind. After thinking about the problem for a long time, I considered many possibilities: maybe this was a book of military tactics written the way it was to disguise the plan of attack. Maybe it was merely a book of drawings, and the artist wasn't writing names for diagrams but was expressing their own response to their art: "This artwork is great!" may be what one of the phrases means, which is why it appears so much across such diverse subject content. Or maybe the phrases are even more general and undefined, like "Big", "Small", "Green", "Yellow", "Three", "Ten", "Good", "Bad", etc., etc..

These were all interesting theories to think about, until there was one idea that seemed most plausible to the evidence, even if it seems involved and complicated otherwise. The stars and the plants are tied, but how? Definitely with grow cycles. Certain plants only grow at certain times of the year. Stated differently, certain plants only grow when certain stars are visible to someone on Earth. So why not hypothesize that the stars themselves are the rain, and that once those stars have "rained themselves out", they disappear from the sky, the plants related to those stars stop growing, and one can no longer see the star? Wouldn't that mean that the plants of that month come from that star? Wouldn't you even be able to say that those plants *are the star?*

In every astrological chart, the naked nymphs are holding stars while sitting in buckets, while in the last chart, the buckets are tipped over -- to no distress at all of the nymphs. On one of the pages of the magical/botanical section (page 153, "f78r"), there is a chain of these bucket, thimble-like seats floating upward, suggesting evaporation, or the implication that the star's seasons are cyclical. This theory is neither supported by nor contradicted by the much-earlier stated "Working Theory," which was based solely on the diagrams. This new theory is based on word analysis of the text. However, it seems so fantastic, that even I am skeptical of it; but, there seems to be strong evidence that this theory, or some variation, may have some truth to it regarding the mythology of the writers of the Voynich Manuscript.

Stage 5: The Corroboration

As of this writing, I have not attempted to write any code that could corroborate the above theory, though I have designed some algorithms that might be of use in either corroboration or furthering the decryption. And furthering the decryption might lead to some corroboration or contradiction of the mythological theory. The first and easiest idea at corroborating the theory is to continue the decryption using the above-mentioned human algorithm: guess a word, check, and either keep guessing or remove the previous guess. Even though it's easy, it would require plenty of time, which led me to thinking about other options.

The most complicated solution would be to develop a linguistics-guessing machine. It would create a hypothesis for grammar organization, such as "[noun] [adjective] [verb]" or "[adjective] [noun] [verb]". Then it would make comparisons of the known nouns (from the Named Diagrams set) with the other words, to determine the only possible types for other words. However, this clearly is outside of the bounds for a mathematician and cryptologist to complete: it requires a heavy understanding of linguistics, which I simply do not have. Learning astronomy took me enough time to understand the book's astrological charts, so I'll leave the linguistics-guessing code as an idea.

There were two final corroboration ideas that seemed to be most efficient to the task and possibly sufficient for verifying the theory. First, there is the Proximity Word Search, which searches for common words that appear near known word types, such as words that appear very close every time to words used to describe stars -- these words might be decrypted as "month", "time of year", "sky", or any of the many other possibilities. This method is much more oriented toward mathematics and computer science.

Second, there is an astronomical and botanical comparison, very different from the above methods, but much more geared towards corroboration. If it's true

that the stars are rain and this means that the plants that grow at certain times are those stars, then we must be able to compare the astronomy to the botany. We know which months astrological charts are talking about by the signs of the Zodiac. Why not associate plants with each astrological chart, according to the words that appear for stars on their pages? Then we would be able to have a calendar based not on stars or days, but on plants as they appear in the wild. This would definitively prove a strong association between the plants and the stars, the closest that I can imagine getting to corroborating the above theory. Of course, it requires far more astronomy and botany knowledge than I possess. And so I leave it, too, unfinished.

Conclusions and Possibilities

"Okeol cheody qolkeey ocheky chody otchdy oko okos." This was supposed to be the 50% decrypted sentence that was going to lead to decrypting the entirety of the Voynich Manuscript. But this turned out not to be the case. Instead, the only theory that could explain all of the accumulated evidence was this fantastic mythology that associated stars with rain, gods, and plants. And even this theory has no definitive corroboration.

There are some who say that the Voynich Manuscript is a hoax. Not only do I know that it is authentic -- but I feel quite capable of identifying the geographic region in which it was written, or at least, in which its author was trained to write. The final decryption analysis never decoded the full meaning of a single word, but it gave some general idea about the thinking within the book. When it comes to a book that has failed to be decrypted after hundreds of years with some of the best minds of cryptanalysis, one should be cautious about giving certain answers. Even though I came up with a theory, there is a lingering disappointment that I was not able to come up with anything more definitive.

Like Voynich, I believe it's fine that someone is communicating meaning to me which I may never understand; I'm not going to be like some scientists who become so tremendously frustrated at this project, that their only conclusion is that it cannot be done or the goal itself is a hoax.