The Voynich Cipher Disk

Heinz-Dieter Hauger

hdhauger@gmail.com

April 24, 2022

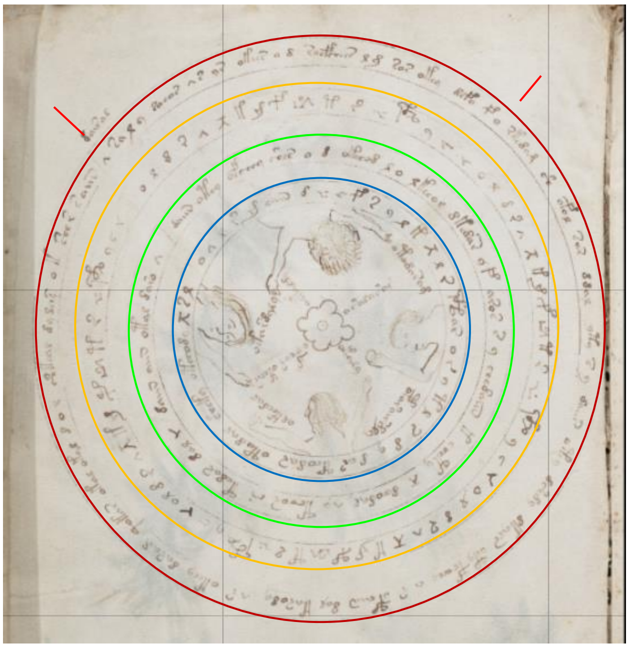
Abstract: The Voynich Manuscript VM is so far not translated nor placed in any context despite numerous attempts made throughout the years. This is owing to an ancient still patentable crypt mechanism that converts classified text into words of syllable-tokens that appear and sound like poems from a long-lost ancient language. The instructions to use this mechanism, illustrated in the VM, and an example is supplied to demonstrate its use. The manuscript's original language is revealed to be Ottoman-Turkish. The Google translation of a transliterated sentence is given and displayed to match the page's overall concept. Finally, the procedure for resolving the entire Manuscript's translation is given.

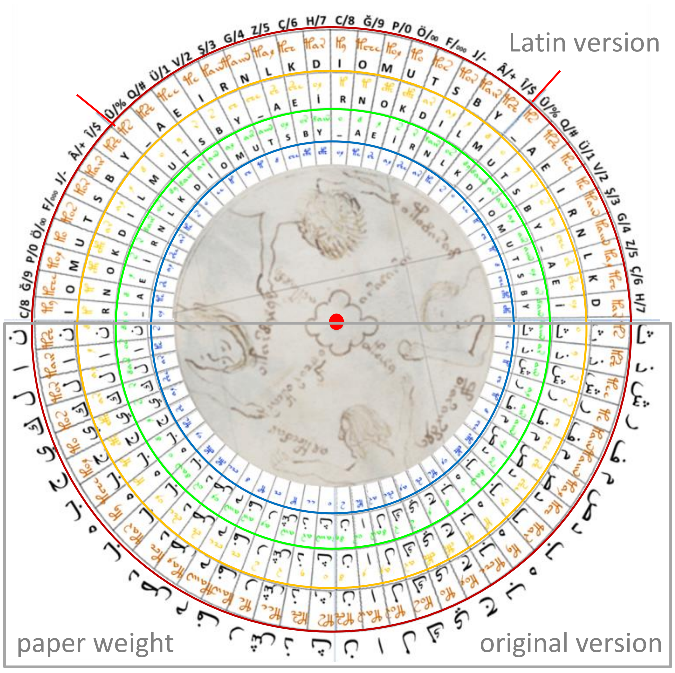
The article takes about 15 minutes to read.

It has always been there

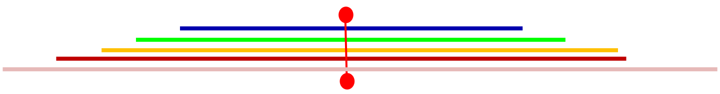
The four center-mounted parchment disks are depicted on page 57v, with the rings for the original letters left blank and only some of the real tokens visible, but the ring with the four times repeated 17 character-sequence gives up the 68 segments required for each disk. The behavior of the person seen in the middle, raising two hands to welcome the user, making half turns, and ultimately holding up the winner's wreath, accurately describes the use of the mechanism:

"**You are welcome to turn the 4 disks to the correct positions and be rewarded by reading the original text**."





Page 57 v *stacked Parchment disks with string or rivet*



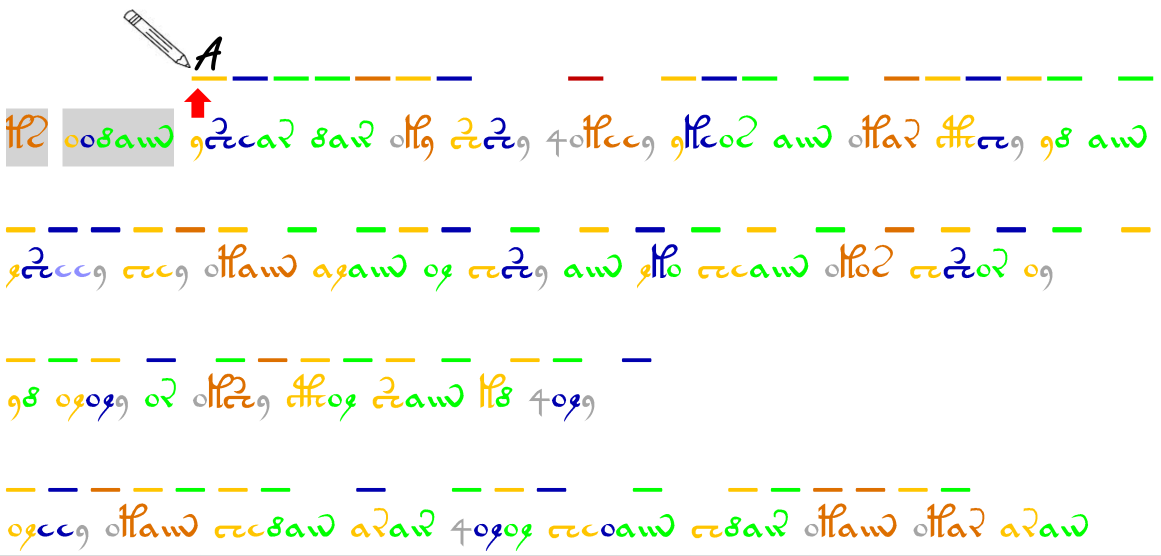
The original disks requires Ottoman letters all around the blank rings, but we switched to Latin letters for the upper half to illustrate the function to those who do not understand Ottoman. The mechanism is shown on the original page with the radial lines that mark the setting tokens (in this case) at a 45-degree angle, allowing the 17 upper segments of the three inner disks, which hold the most frequently used letters (87%), to be read comfortably while the 34 segments on the upper outer disk only require a slight inclination of the head towards the shoulders. To avoid misalignment after the code was set, the lower part was probably forced down and covered by a paper weight.

Each word in the Manuscript, such as , must be broken down into INITIAL, MEDIAL and FINAL tokens of one or more characters in order to be found on the appropriate disks. The most frequent letters are represented by the three inner disk's 51 tokens (repeated 4 times), the less frequent letters by the 34 tokens on the outer disk .. (repeated twice), and Digits are denoted by the grey prefix .. alternately found on the outer disk (= , , or). The dials are set for the following chapter to four not translated tokens inserted at the beginning of the line on all encoded pages. If or characters are enlarged and the end stroke made pointing left to the beginning of the line (or), the dials must be re-set as described above. Grey characters such as are placeholders that allow a token to be placed in the desired position within a word( ). The Final tokens may stand alone. **Spaces between VM words** are part of the tokens not to be mistaken with spaces "_" of the original text, which are encoded like letters. We included special letters for different languages just like the Ottoman alphabet had Persian, Arabic and Urdu letters.

Try it out

The following text is **not** from the VM but it appears to be, because each of the token-words occurs multiple times in the VM, of cause with different unknown meaning. Translate this newly created text to proof that information **can** be hidden inside the VM. To solve the example below the table is sufficient, but if you wish to encode your own text, replicate the linear table horizontally on one piece of paper, cut along the colored lines then move the paper strips against each other to set your code. For round versions, duplicate the Latin half, copy five times, cut out four rings of increasing size, then stack and tack them centered. Claim your prize by using a mechanism hidden for 500 years.
Find the colored tokens in the table and scribble the black letter above the colored dashes with a pencil.

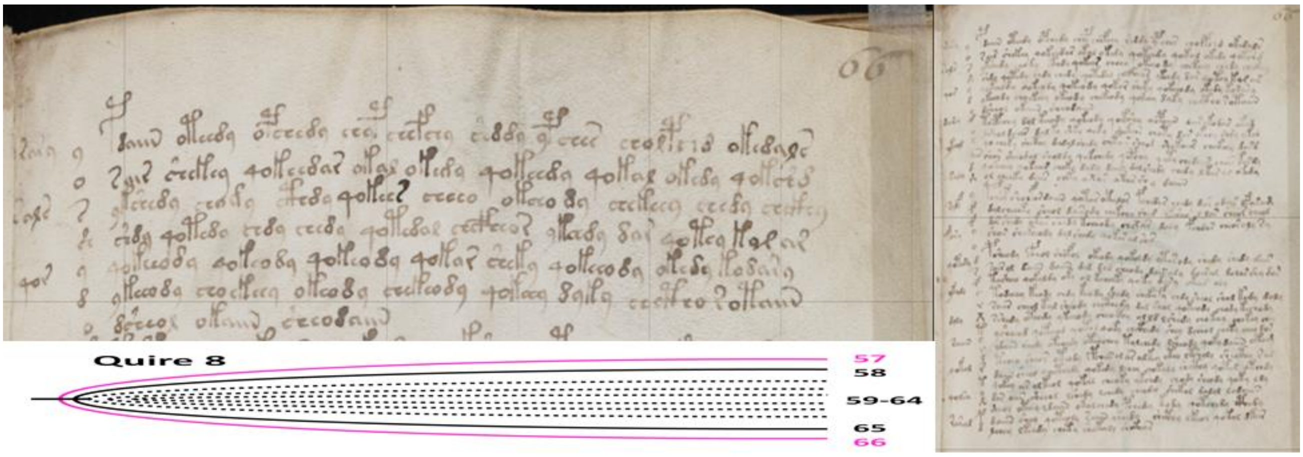
e.g. = A = 1 Do not translate grey letters . Takes only 3 minutes



/	Û/%	Q/#	Ü/1	V/2	$/3	G/4	Z/5	Ç/6	H/7	C/8	Ğ/9	P/0	Ö/00	F/000	J/-	Â/+	Î/$
outer																	
	_	A	E	İ	R	N	L	K	D	I	O	M	U	T	S	B	Y
initial																	
	U	T	S	B	Y	_	A	E	İ	R	N	O	K	D	I	L	M
final																	
	D	I	O	M	U	T	S	B	Y	_	A	E	İ	R	N	L	K
medial																	

***Note:** The table shows only ¼ of the disk and is set to the 4 tokens of the chapter start (compare).* Verify your findings here.
Repeat the transliteration in **black and white** to verify if you understood the method to identify the tokens in the VM.

Detailed Instructions



Page 66r also in the instruction **Quire 8** on the **same** parchment as 57v explains the rules and exceptions for the token words and characters needed to produce precise translations. The 16 words in the front column are special examples of coded words; their tokenization is explained in detail in the coded text to the right. The second column of single characters is needed to explain the pronunciation of the VM characters to Ottomans as explained later. Please observe that about every 3 lines the and mark the lines where the settings of the disks must be changed. Read the following lines only if you are interested in the coding details.

initial medial with untranslated as ending to distinguish from

initial with medial and as final

untranslated with medial and final (not exception)

initial medial terminated by which is same as (with failed down stroke)

initial medial with visible connection line and as final

3 initials terminated by can precisely looked up in the table

cannot be finals so is clearly a medial and no final is needed

read as but marking a new chapter

number with added medial token (all and require phonetically an ending)

2 medial tokens in a row and

initial and final

number normal no needed since is an end syllable like all finals are

now is a 2 character medial token not

followed by , , , , at the word end are always finals

All of **Quire 8** is probably encoded with a different disk and used identical in other books and documents that might be found in the "Office of the Prime Minister Ottoman Archives". This permits an instructor to train the end user on how to use the mechanism without revealing the rest of the classified information. The end user had two disks: one that he shared with the teacher to study and one that was exclusively for him.

The enormous amount of work that went into the book, as well as the particular range of topics covered, lead to the conclusion that the VM was left by the sultan himself in the form of a time capsule to instruct future generations of sultans.

The Evidence

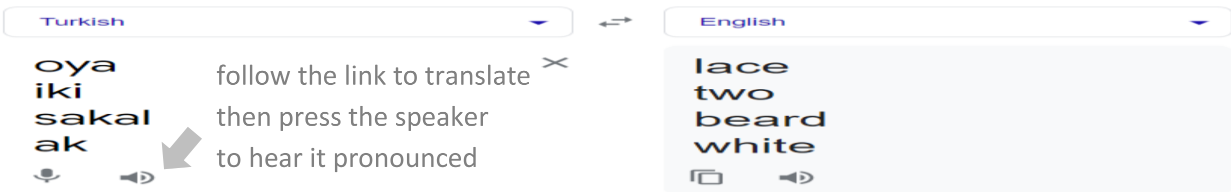
While most of the VM is written in coded chapters, many stand alone words not related to any chapter could not be encoded because the four tokens needed to set the Cipher Disks can not be associated. In this case the VM characters (not tokens) are just phonetically transcribed and some resulting words may still be understood by native Turkish speakers. In 2019 Ahmet Ardiç's ATA Team published a transcription alphabet to Latin-Turkish, discovered many words to correspond to illustrations and demonstrated that the Voynich characters were once recognized in the region. Mr. Ardiç is capable of pronounce-reading Voynich text following the Turkish language rules (the ATA alphabet pronounces some letters differently).

The author has no knowledge of Turkish or Ottoman but found this convincing example:

Page 65r shows a **cut back** male cannabis farm plant with typical 5-leaved flowers and seeds. Male plants are usually **rooted out** because female plants are of better quality. At an early stage of growth, when this task has to be done, their gender is difficult to distinguish.

On the left side of the plant the VM text  is written and pronounces **oya ik sakal ak**.

Google™ translates this from Turkish to English as: " **Two lace of white beard** "



Wikipedia™ informs that this distinguishes the female plant at the early stage of growth (item 6).

Female plants develop pre-flowers resembling a pear-shaped ball from which two small, usually white hairs grow.

The image on the page shows the outrooted flowering male plant when easy to distinguish by its seeds, while the written information explains the best and earliest point in the grow phase to weed out. If a detailed drawing of a natural plant and a sophisticated abstract concept coincide with four words in a tested language, the likelihood of the language being correct is almost certain, but there is still more evidence pointing to Ottoman-Turkish that was used for 600 years until 1928 when it was replaced by Latin-Turkish.

The Ottoman

The map on the VM Ros page depicts the corners of the globe as well as the trade routes, which were top secret at that time. In the middle minarets of Islamic mosques erected and praised only by Ottoman sultans are shown. The roof shapes were inspired by the era's turban trend. At the time (~ 1420) the WM was dated to, the Ottoman Empire was the only Islamic state. Ottoman script was used in millions of books, journals, chronicles etc. The Başbakanlık Osmanlı Arşivleri ("Office of the Prime Minister Ottoman Archives") in Istanbul still keeps the central State archives, with around 150 million documents of which 60 million are classified (Özkul, 2000). This highlights the need for encrypted communication within the vast Empire. Several Ottoman characters are shown on 57v and 66r, with letter و Waw (w, v, o, u) used in the VM 17,000 times in same positions at beginning and end of words like in Ottoman or todays Arabic. Women are depicted on many pages with for that time unique Ottoman fore-head ornaments and hairstyle.

Ottoman script is similar to Arabic, but has 33 letters. Single words can be phonetically represented by Latin letters but in sentences words agglutinate without spaces between, so to discriminate the words and grammar the position of a letter in a word (STAND ALONE, FINAL, MEDIAL or INITIAL) is paramount. Most letters therefore have four different forms of writing(ínwinteriski find and discriminate i i i i ی ـی ـیـ یـ the 4 forms of Ottoman "i").

A Disk Re-engineered

A statistical analysis of the VM detected the proposed 119 tokens and related them to their functions. The VM is composed of only 78,000 token each standing for a coded Ottoman letter. 68,000 (87%) are of the most frequent 51, the remaining 68 are less frequent letters and numbers located on the outer disk represented by … or … words. Since almost all words end in one of only 17 final tokens (syllables) the text looks and sounds like a poem. Tokens are carefully chosen to confuse recognition and built up well sounding token-words that follow Turkish syllable rules, therefore the encoding has remained undetected so far. Because reading letters upside down on a disk is difficult, only the upper half can be used, so the lower half must be an exact copy. Of course, the secrets in text and images were unknown to the artists who made the VM. The text was encoded in advance by an agent with security clearance on a piece of paper and after this the Cipher disk was never kept together with the Manuscript, as credit card pins are kept nowadays. Pictures were copied from other documents to the parchments however the artists might have never ever seen similar plants before. Line ends or branches may have required hyphenation of lengthier words while a calligraphist copied the encoded text to parchment. In this case the word can be interrupted by and continued by , or in the next line or on the opposite side of the branches without the need of re-coding the whole word which for the artist was impossible to do. If the original text has rhymes, the tokens show a diagonally displaced repetition, because the rhyming syllables would not align anymore, because the coded text is 2.5 times longer(/ page 15v).

Normally, prose does not include many numbers and signs (4o..) , unless historic events are listed as seen from page 103r onwards, where stars are employed as bullets to mark successive but unconnected events, a method that is still in use. Many of these patterns produced by the encryption can be found in the VM. You can easily search for them once you downloaded the MS-Word Voynich file.

It has been known since the Middle Ages that a table encoded text with more than 20 letters is trivial to crack using frequency analysis. Using four tables, the coded chapters needed therefore to be terminated after maximal 80 letters and a new chapter be started with different settings, until the original chapter ended. The coders made sure that the four setting tokens in the VM's 2000 chapters never repeated, most likely using a checklist. For convenience in consecutive chapters of the same page they often only changed one of the four setting tokens turning one disk. The three inner disks had a total of 17 × 17 x 17 = 4913 possible settings, therefore no two chapters had to be encoded in the same way. Otherwise broken up words can only be distinguished, if all three inner disks are set accurately. Redundant safety was applied: the characters were exotic, the encoding disguised, and the original language unknown. Even those who knew all that, such as other agents with security clearance, would have to permute each chapter for 5000 times 5 minutes to see the first words, and even then, the precise content can only be known if the original disk is completely reconstructed, which is impossible to do manually in a lifetime.

Today still a difficult but possible Task

Knowing that the VM's original Ottoman text was transliterated using tables, chapters with a large number of repeated tokens might be attacked by brute force, sorting out transliterations with impossible bigrams and trigrams. The enduring are compared to Ottoman dictionaries to identify words. A probability is created for each permutation based on the quantity and length of words detected. The 34 letters on the blank rings and their relationship to the tokens may be readily identified this way. The 68 tokens of the outer ring must be permuted with the remaining Ottoman letters and used in the chapters with the highest word count that have already been found. The same selection process as before will produce the remaining letters on the outer ring and identify more words. Finally, the chapters with most found words can be linked to the setting tokens hence allowing accurate transliterations of all encoded chapters into Ottoman script. All of this may be accomplished with the help of a computer and no prior knowledge of the language. For unknown reasons, Google does not translate Ottoman, but Miletos.co and certain scholars can translate Ottoman to Turkish and from there to English would be a breeze.

A MS-Word file of the whole VM is available that presents the characters as the original VM and allows you to compare, analyze, color, highlight and search the text in the same way you would any other document. With the corresponding font Excel and PPT can be used, too. The author readily provides any reasonable assistance to solve the translation.

The suspicious word on the upper left side of 57v 8airog pronounces "**şair ol**" and translates teasingly to**: be a poet**...

References:

1 The invaluable " Voynich Manuscript Voyager" Jason Davies (2017)
https://www.jasondavies.https://www.jasondavies.com/voynich/#f57v/0.866/0.425/2.75com/voynich/#f57v/0.866/0.425/2.75

2 "The Voynich Revisionist" Julian Bunn (2022)
https://voynichrevisionist.com/

3 "The Cardan grille approach to the Voynich MS taken to the next level" René Zandbergen (April 2021)
https://arxiv.org/ftp/arxiv/papers/2104/2104.12548.pdf

4 " Ottoman Turkish alphabet " Wikipedia (April 2022)
https://en.wikipedia.org/wiki/Ottoman_Turkish_alphabet

5 " Ahmet Ardiç's ATA Team" Ahmet Ardıç (April 2022)
http://www.turkicresearch.com/

6 " Machine-Based Transliterate of Ottoman to Latin-Based Script" Ashti Afrasiyaw Jaf (November 2021) Researchgate
https://www.researchgate.net/publication/356162643_Machine-Based_Transliterate_of_Ottoman_to_Latin-Based_Script

7 "A deep learning based translation system from Ottoman Turkish to modern Turkish" Abdullah Bakirci (2019) Gebze University
https://acikbilim.yok.gov.tr/bitstream/handle/20.500.12812/268300/yokAcikBilim_10297160.pdf?sequence=-1&isAllowed=y

8 " Statistical digram and trigram analysis of Turkish ..." Baran Uslu et al. (Jan 2010)
https://www.researchgate.net/publication/228821437_Statistical_Digram_and_Trigram_analysis_of_Turkish_in_terms_of_coverage_and_entropy_for_possible_language_and_speech_based_applications

8 " Ottoman OCR System" Miletos (2022)

http://miletos.co/en/showcase/ottoman-ocr