Benedek Láng

Why don't we decipher an outdated cipher system?
The Codex of Rohonc[1]

A neglected source

The Codex of Rohonc is in a paradoxical situation: it is privileged to be in the elegant company of the nine most famous hitherto unsolved writing-systems (together with the Phaistos disk and the Linear A),[2] while – in contrast to all the other unsolved writings – almost no serious research has been published on it since it turned up in 1838.[3]

This situation becomes particularly surprising if we compare it with that of the (in)famous Voynich manuscript. There are a couple of similarities between the two manuscripts: both have been written in a mysterious and unsolved cipher, both are fairly lengthy (and thus offer themselves to professional code breakers), both can be equally well a cipher, an artificial (or perfect) language or a hoax, both of them contain illustrations that might help the process of deciphering or might equally well misguide us, and both could have been written – or forged – sometime between the 16th and the 19th centuries. However, while the Voynich manuscript enjoyed the attention of the best code breakers of World War II, many twentieth century philologists researched it, and it has been offered several monographs, articles, web pages and e-mail lists, the Codex of Rohonc has only received a few aborted attempts at deciphering it and one meager e-mail list. There is not even a reliable piece of secondary literature on what can be actually known about it.

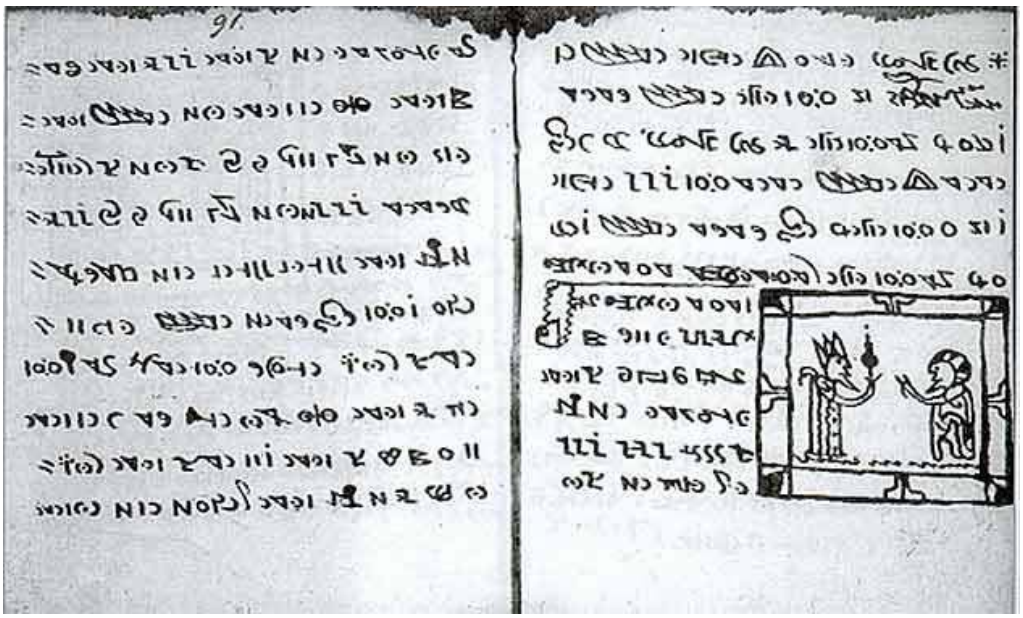
The Codex of Rohonc is a nearly 450-page long handwritten paper book filled with 9-14 lines of unknown sign-strings on each page and with more than 80 illustrations. Nothing is known about the provenience of the manuscript, even though it is today in Budapest, its Hungarian, or even East-Central-European origin is uncertain. It was donated to the Library of the Hungarian Academy of Sciences together with the 30 000-volume library of the late Hungarian count, Gustav Batthyány in 1838. This library was earlier located at the family residence in the town of Rohonc (today: Rechnitz, Austria), hence the name of the codex. The

[1] My research was generously supported by the Mellon fellowship in the Maison des Sciences de l'Homme, Paris, and also by two Hungarian Grants, the OTKA K 72598, and the Magyary Postdoctoral fellowship. I would also like to acknowledge the intellectual help of Dóra Bobory, Gergely Buzás, Ottó Gecser, Eduard Frunzeanu, Cristian Gaspar, István Monok, Veronika Novák, Dóra Sallay, and Márta Tarnai.

[2] Cf: http://www.omniglot.com/writing/undeciphered.htm

[3] The codex is kept today in the Library of the Hungarian Academy of Sciences: MS K 114, or MF 1173/II. The Wikipedia article on the codex provides a fairly reliable overview: http://en.wikipedia.org/wiki/Rohonc_Codex The pages of the codex can be downloaded from this web address: http://www.dacia.org/codex/original/original.html

initial enthusiasm of 19th century Hungarian scholars for the supposedly early-Hungarian script was soon followed by disappointment, skepticism and suspicion, and late 19th century scholarship came to the conclusion that the codex is definitely not a Hungarian linguistic monument, but rather a forgery.[4] Since then, serious historians and philologists avoid dealing with it, or more precisely: they do not publish about it.



Two pages of the codex of Rohonc (91-91a)

Current state of research

Earlier research on the codex was carried out mostly by self appointed historians and amateur code breakers. This is not necessarily a problem, since several undeciphered scripts and mysterious ciphers have been solved by amateurs,[5] in this particular case, however, the lack of historians and philologists did not help the professionalization of the research. The codex surfaced for different kinds of projections: it has been suggested that it contains an old-Hungarian,[6] a proto-Rumanian (11th century „vulgar Latin"),[7] or an old Sanskrit script.[8] As I argue extensively in a recent article of mine,[9] – and as other critics have also pointed out one

[4] The acts of the Linguistic Committee of the Hungarian Academy, 12th Nov. 1898. MS Ral K 1568.
[5] There are many such examples reported in the articles of *Cryptologia*.
[6] See two attempts: Kálmán Némäti, *Rohonczi Codex Tantétel* (The Codex of Rohonc Doctrine), Budapest, 1892 (he does not give a solution, just claims that the codex was written in Hungarian); and Attila Nyíri, „Megszólal 150 év után a Rohonci-kódex?" (Does the Codex of Rohonc Start to Speak After 150 Years?), *Theologiai Szemle,* 39 (1996), 91-98.
[7] Viorica Enăchiuc, *Rohonczi Codex: descifrare, transcriere si traducere*, Bucarest: Editura Alcor, 2002.
[8] Makesh Kumar Singh, „Rohonci Kódex" (The Codex of Rohonc), *Turán*, *2004/6 = 2005/1*, 9-40.
[9] Benedek Láng, "Miért nem fejtünk meg egy elavult titkosírást? – 1 (A „rohonci kódex")", (Why don't we decipher an outdated cipher? – 1 The codex of rohonc) in *"Mielz valt mesure que ne fait estultie". In honour of*

by one[10] – there are some problems with these reconstructions. First of all: none of the code breakers give their correspondence-table, that is, a list on which the reader can check which sign corresponds to which letter or syllable or word or notion. And if the readers try to prepare this table themselves, they will find that different letters of the deciphered texts correspond to the same signs of the codex, and vice versa. This would be, of course, possible in a polyalphabetic cipher system, but there is no sign that the code breakers would suggest such a system, they rather believe that the sign strings of the Codex of Rohonc are the letters of a natural language. However, corresponding different signs of the cipher to different letters of the plain text without any systematic method is on such a level of arbitrariness, where every undicephered script can be solved with any meaning.

Three of these attempts do not even recognize the direction in which the signs are to be read. Although we cannot decipher the series of signs, and we cannot even be sure whether they incorporate a real system, or only a hoax, the sign strings decidedly have a direction. The lines are justified to the right, on the left side of the pages we often see hyphens, when the text is finished, there is free space on the left hand side and on the lower part of the page, and finally, when a longer series of signs is repeated, and has to be broken because of the end of the line, it is always continued on the right end of the lower line. So interestingly, while we do not know what it means, and whether it means anything, the text has a direction, and it is: from above to the bottom, from the right to the left. And this is what three of the solvers (Attila Nyíri, Viorica Enăchiuc, and Makesh Kumar Singh) do not take into consideration – even though many earlier scholars, who did not actually claim to have deciphered the codex, rightly identified the direction.

That prior beliefs can seriously influence sheer perception is demonstrated by the Rumanian archaeologist, Viorica Enăchiuc, who identifies the heroes of early Rumanian history in the illustrations, where everybody else sees Biblical and other Christian scenes (such as the crucifixion, the annunciation, the three magi in front of the infant Christ (with the star of Bethlehem on the picture), the entry to Jerusalem, Christ in front of Pilate), that is, scenes having a fairly standard and recognizable iconography.

Two more attempts should be mentioned, not because they were more successful, but because they were more honest not claiming to have solved the text, and because they

---

Iván Horváth. Eds. István Bartók, Béla Hegedűs, Mihály Szegedy-Maszák, Márton Szentpéteri, Levente Seláf, András Veres, 141-147. Krónika Nova Kiadó, Budapest, 2008.

[10] Ottó Gyürk, „Megszólal a Rohonci-kódex?" (Does the Codex of Rohonc Really Speak?), *Theologiai Szemle,* 39 (1996), 380-381. Dan Ungureanu, „Nu trageti in ambulanta," *Observator Cultural, 167* (2003). Csaba Varga, „A Rohonczi Kódex M K Singh-féle olvasatának ellenőrzése" (A Criticism of M K Singh's Deciphering of the Codex of Rohonc. A letter to the editor), *Turán* 2005/2-3, 198-202.

followed a more professional agenda. Ottó Gyürk and Miklós Locsmándi tried statistical methods to recognize the patterns according to which the signs tend to follow each other.[11] None of them believe that the codex was written in a natural language; rather both are inclined to think that the language system is a constructed one.

As a result of the past 170 years, our knowledge on the Codex of Rohonc is quite incomplete. Thanks to the watermark, we do know that the script was written on paper produced in the 16th century, we do know that – if in any direction – the text should be read from right to left, and we do know that its illustrations concern mostly Christ's life, crucifixion and resurrection. However, we do not really know, whether the content of the codex is really religious, we could not even compile a finite list of the signs used (there are some 150 different signs in the codex, but it is always possible to find new ones), we do not know whether it is a cipher or a code system, whether these are syllables, letters, consonants or full words that are substituted by the signs, and finally we do not know which language (or even which geographical area) can be identified behind this system.

Starting points

Finding analogies among ciphers, perfect languages, shorthand systems, and hoaxes and understanding how such systems were constructed may provide hints as to how the coding system of the Codex of Rohonc really works. In addition, finding analogies among religious texts for the illustrations and for the structure of the text may provide so called „cribs", words or notions that probably occur in the text, finding which may help decode the signs, and determine whether signs stand for letter, syllables or words. Finding the very language that was decoded is unfortunately a step that can be made just after answering at least some of the previous questions. Even though a satisfying solution of the sign strings of the codex is not yet available, it is still achievable to narrow down the possibilities of who, where, when, and for what purpose produced it.

The question of forgery

Late 19th century Hungarian scholarship became convinced that the codex was forged by the antiquarian Samuel Literati Nemes (1794-1842), who provided many aristocratic families with old books, but also with a number of forged Hungarian "linguistic monuments", among

[11] Ottó Gyürk, „Megfejthető-e a Rohonci-kódex?" (Can the Codex of Rohonc Be Solved?) , *Élet és Tudomány* 25 (1970), 1923-1924; Miklós Locsmándy, „A Rohonci Kódex. Egy rejtélyes középkori írás megfejtési kísérlete," (The Codex of Rohonc: An Attempt to Decipher a Mysterious Medieval Script) *Turán*, 2004/6 = 2005/1, 41-58.

them historical maps, a pseudo Chinese writing and a number of other charters or illustrated book pages that first seemed to be many centuries old but that were ultimately debunked by philologists. Literati's forgeries have been gathered together in two folders (National Széchényi Library, Fol. Hung. 1365). Two things are common in these pseudo antiquities. First, they are usually relatively short, while they may contain many illustrations (of rather low quality, actually) only short texts appear in them. Second, they all claim to be something (usually Hungarian linguistic monuments), which they are not (that is why, we call them forgeries).

Does the Codex of Rohonc really belong to this infamous company? While in the case of most of these forgeries we have external evidence that they come from Literati, in the case of the codex nothing supports that, and what is more, comparing the chronology of Literati's life with the emergence of the codex, this is not very likely. Second, the codex is very long compared to what Literati preferred to forge: why would a forger choose to prepare such a long (almost 450 page long) codex to deceive his contemporaries, when a few dozen pages would be more convenient for this purpose (not least, because shorter texts can be deciphered with greater difficulties, or cannot be deciphered at all due to the lack of sufficient information)? Third, the Codex of Rohonc does not try to sell itself as an old Hungarian text, as a matter of fact, it does not seem to contain a medieval or an early modern natural language (there are no spaces between the words, the signs are not "letter like" and there are too many of them, and it is not clear what purports to be a letter and what a word, etc), it was only the 19th century Hungarian literary community that saw old-Hungarian texts everywhere (and many times they were right, sometimes deceived).

But if not Literati, someone else still could have forged the codex, perhaps not with the intention of making it seem old Hungarian, but rather with the intention of selling it as a mysterious and precious book. Some of the Voynich scholars believe that this is exactly what was behind the production of the Voynich manuscript. However, while the Voynich manuscript is a beautiful codex with illustrations of genuinely mysterious content, the Codex of Rohonc is not particularly beautiful, its illustrations do not promise interesting content, only pure Biblical matters (but not even an apocryphal gospel). A further problem is that the structure of the sign strings is quite strong, long repetitions appear in the codex far from each other even in sections inscribed by different hands. So, if someone (or more precisely: some ones) decided to prepare a hoax, they had to agree previously in the fairly complicated set of "grammatical rules" of the pseudo-text.

In spite of the counter arguments, I cannot rule out the possibility, that the Codex of Rohonc is a hoax, but I would refrain from using the word forgery, because it implies that we know what it purports to be while it is not that in fact. I believe that if it is a hoax indeed, it wishes to sell itself not as a text written in a natural language but rather as a cipher or a perfect language. However, if this is the case, our task is to understand why someone chose to prepare a pseudo-cipher or a pseudo perfect language in the 16th-17th centuries, a question that brings us to the same archives as a research on real 16th-17th century ciphers and perfect languages.[12]

The paper

As the watermarks let us delimitate the production of the codex, the paper was made in north Italy, in the years 1530-1540.[13] As a rule, it is quite unlikely that a paper codex remains unwritten three centuries after its creation, because such codices were fairly expensive. It is more likely that the codex was written not much later than its paper was produced, that is, in the late 16th or in the early 17th century. However, we cannot completely exclude the theory of a later, even 19th century inscription, since unwritten codices – rarely – turned up in bigger libraries, as was the case for example in the extensive book collection of the Battyhány family,[14] from where the Codex of Rohonc emerged. All in all, we have a *terminus ante quem*: 1838, and a *terminus post quem*: 1530 for the inscription of the codex.

The illustrations

Although the illustrations are very simple and even naïve, their careful study can help identify the genre of the script, and also date the codex.

As it has been mentioned, a great part of the more than 80 illustrations are clearly biblical, they depict scenes from the gospels, such as the crucifixion, the annunciation, the three magi in front of the infant Christ (with the star of Bethlehem on the picture), the entry to Jerusalem, Christ in front of Pilate, the crucifixion, the "*noli me tangere*" scene, the

[12] The alternative that the codex might be a forgery in the sense that it contains an invented pseudo-language (similarly to the famous case of Psalmanazar) will be treated in the section on invented artificial languages below, because it raises the same kind of questions as invented languages that were constructed without the intention of deceiving others.

[13] See Charles-Moïse Briquet, *Les filigranes, dictionnaire historique des marques du papier des leurs apparitions vers 1282 jusqu'en 1600.* Paris: Picard, 1907, 507-508. Since very similar watermarks were used in north Italy in the period 1530-1600 (see Briquet 477-532), the exact identification depends on how precisely we reconstruct the 'anchor with the star' that is the form of the watermark. According to Briquet, this specific form refers to the participation of Venetian masters in the production of the paper.

[14] András Koltai, Batthyány Ádám és könyvtára (Adam Batthyány and his library), (Budapest-Szeged: OSZK-Scriptum, 2002), 256.

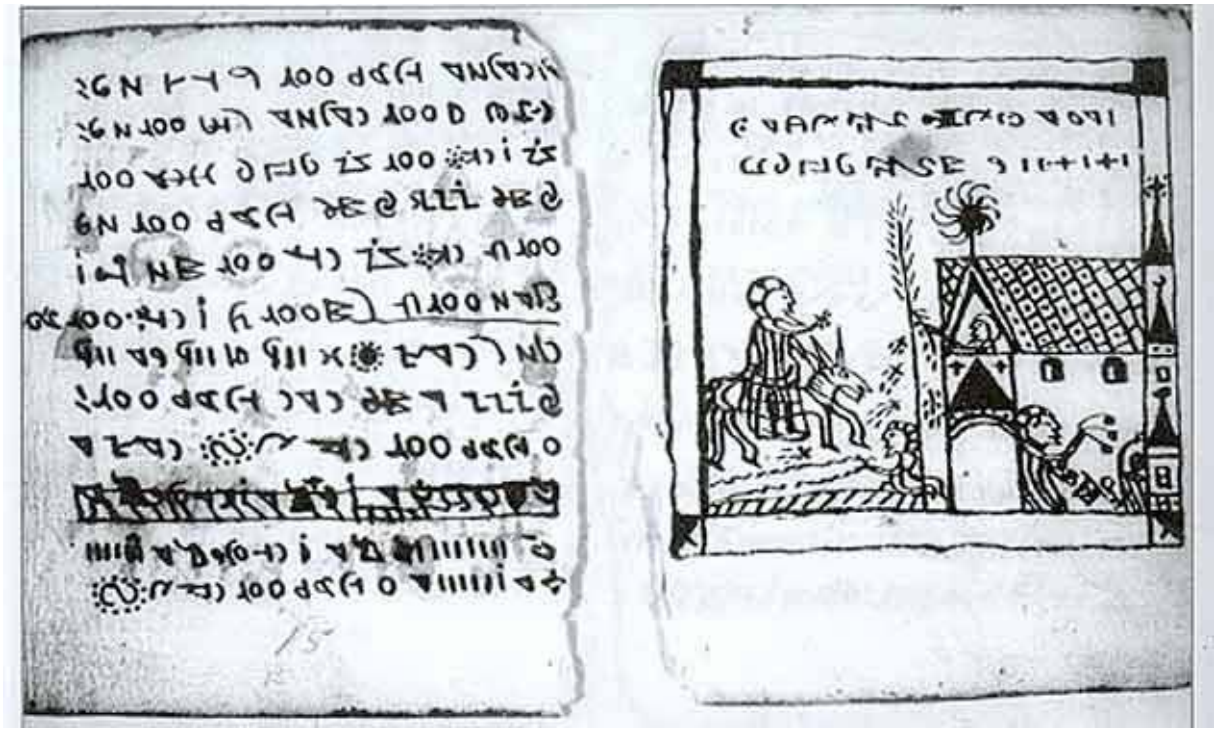
resurrection, etc. Later images (such as the one reproduced on our first picture above) are not that easy to identify, as they depict usually two persons – one of whom seems to be usually Jesus – involved in conversation. These images might refer to specific parables, others to the scenes from the Apocalypse of St John.[15]

While the images are simple and sketchy, they betray some information. Many of the elements appearing on them (misunderstood gothic features, the shape of the weapons, the structure of the churches, etc) indicate the 16th-17th centuries, and (particularly the shape of the buildings) do not show beyond the eastern border of Central Europe. This means that the codex might be of Rumanian, Hungarian, or Balcan origin, it might just as well be from any western European country, but it is probably not from the Near East – a conclusion that somewhat narrows down the possible languages of the codex.[16]

These images might be also useful for the code breaker in the sense that they provide "cribs", sign strings the meaning of which can be identified and that consequently help the process of deciphering. One of these is the four-letter inscription on the cross, that should be INRI, of course, however, in this specific case, the first and the last letters differ. Also, there is a composite sign appearing in very high frequency in the codex, that probably stands for the name of Jesus, because it occurs often above Christ's figure on the images, among them, on the image where he stands in front of Pilate (p. 41a). The other double sign above the other figure on the same image stands very likely for the name of Pilate, that is confirmed by the fact that this sign occurs very often on the pages around this picture, but nowhere else in the codex. A fourth crib might be the last sign of the letters INRI on the cross, that appears several times above a city on the images, and that might stand for J, the initial of Jerusalem, and that of '*judeorum*' (p. 65).

[15] For the pages of the codex not reproduced here, see http://www.dacia.org/codex/original/original.html
[16] For these considerations, I am grateful to the art historians Dóra Sallay and Gergely Buzás.

The three magi in front of the infant Christ (p. 21-21a)



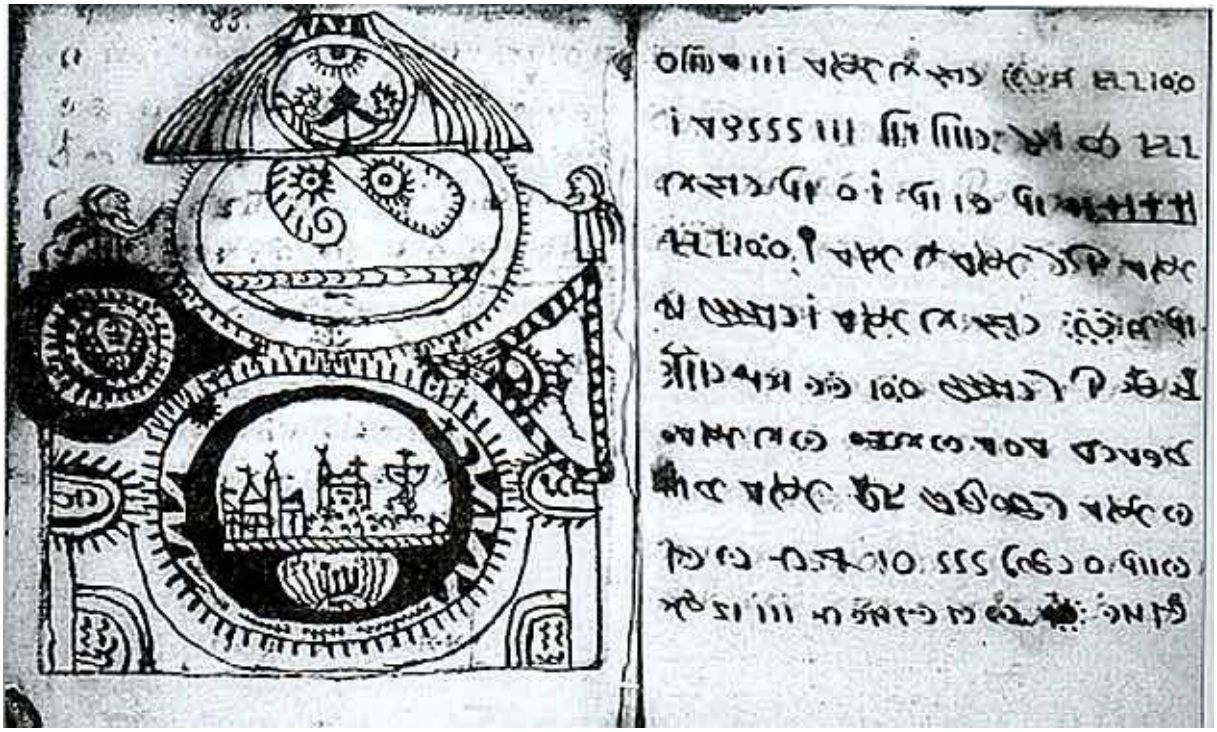
Entry to Jerusalem and – probably – Jesus expels the traffickers from the Church (p. 15-15a)

Jesus looking at Jerusalem (p. 65-65a)

Jesus in front of Pilate (p. 41-41a)

Crucifixion (p. 26-26a)

Apocalypse scene to the left, and – probably – the tree of knowledge, the four rivers, and Jesus, standing on the Chalice, surrounded by two angels to the right (p. 79-79a)



The structure of the word with the spheres and the celestial bodies (p. 83-83a)

These are exactly these signs that let us suppose that the text is not identical with that of the gospels. On the pages that supposedly describe Christ's conversation with Pilate, the names of both speakers appear too many times, thus it seems that this text provides a somewhat more detailed account on the life of Christ. An exciting conclusion would be to suppose that the text is a secret (apocryphal) gospel. While this cannot be excluded of course, it should be emphasized that none of the known apocryphal gospels contain a so detailed account on the encounter of Christ with Pilate, and also that it is quite unlikely that an otherwise unknown gospel survived only in the 17th century source that is written in a cipher or a code language. It seems that we should be satisfied with other, less exciting genres, a *Vita Christi*, a book of hours, a breviary, or a secret prayer book of a slightly heretic sect (not very heretic though, because the images do not show any stronger dualism or other heretic view).

The last identifications would also account for a strange characteristic of the text. Though it is not possible to decipher the signs at the moment, one might notice that they have a very strong repetitive structure. This repetitiveness stands in puzzling contrast with the otherwise great number of signs, which counts more than 150, many of which seem to be composite signs of two or three elements. That characteristic might be caused by the structure of a repetitive liturgical text, a prayer, for example. This supposition is also supported by the

fact that many paragraphs of the text start with the same sign-string (word?), and also ends with it (the one that appears on p. 15a and also on p. 65).

Needless to say, I spent a considerable number of working hours trying to identify the obvious structures and repetitions of the cipher text with known biblical texts, apocryphal gospels (among them late antique Gnostic, and medieval Cathar and Bogomil sources), various *Vitae Christi*, Bible commentaries, books of hours, prayers, and so on, but without considerable success. In a parallel way, I tried to find pictorial analogies in several on-line and printed image collections for the drawings (primarily for those that contain an inscription), and looked through a considerable collection of western European, Christian, Greek-Catholic, and Old Church Slavonic illuminations, wall paintings, drawings, etc, until now, however, I cannot report of any promising finding.

The script

If the text of the Codex of Rohonc is not a hoax, than it must be a consciously encoded or enciphered text. In theory, it can be either (1) a cipher, or (2) a shorthand, or (3) an artificial language.

If it is an enciphered text, then (1a) characters can stand for letters or syllables or consonants, or it might be a (1b) system using both the kind of characters listed above and code words (so called nomenclators) not to mention „nullities“ (such as homophonic systems so widespread in the early modern era did), or it might consist of (1c) codes entirely (and then, strictly speaking, we do not call it a cipher, but a code. If it is shorthand, then interestingly the same three options arise, as shorthand systems were usually composed of characters standing for letters, for letter combinations, and for specific words. Which kinds of characters are dominant – depends on the system. And finally, artificial languages, so popular in the 16th-17th centuries are kind of code systems where signs stand for words, notions, ideas and grammatical constructions. As far as its decoding is concerned, it is similar to code systems listed above as (1c). While all these are – at least in theory – solvable systems, the case is definitely not easy.

In the followings, all these possibilities will be assessed one by one, but it might be helpful to call the reader's attention already at the outset, that regardless the initial motif of the designers of the system of the Codex of Rohonc, regardless whether they wanted to make this text secret, to construct a speed writing, or to create an artificial language in which the Christian truth can be perfectly expressed, our task will be ultimately the same kind of cryptanalysis or code-breaking.

Ciphers

My research done thus far has excluded that the cipher of the Codex of Rohonc would be one of the *simpler* monoalphabetic or homophonic systems. Both frequency analysis and vowel identification[17] strategies failed, and a careful analysis of the text lets me think that signs (at least a great number of the signs) do not stand for simple letters, but rather for bigger particles of language. Polyalphabetic systems are also excluded because of the great number of consequent repetitions in the text, and the obviously low entropy of the signs. If anything, this system should have been constructed according to one of the many methods designed in the late 16th and the 17th centuries, when the vulnerability of simple monoalphabetic systems was already a commonplace, and when polyalphabetic systems were already invented but were not yet applied in practice. In this period, sufficiently complicated homophonic systems were regarded as fairly safe and relatively quick to master and employ. In homophonic systems letters have several cipher equivalents (homophones), frequent characters (vowels) having usually more corresponding signs than infrequent characters, thus the system resists brute force frequency analysis. Nullities, signs without meaning are meant to serve to confuse the code breaker, and a list of signs (nomenclators) stand for specific names, territories and notions, that are often used, in order to avoid to provide the cryptanalysts with cribs, words that he knows to occur in the text, and the pattern of which he easily finds in the cipher. Many of such systems include signs for syllables.

An efficient strategy to understand the system of the Codex of Rohonc would be to find analogies in the famous 16-17th century summaries on stenography and cryptography, as well as among the various political applications of ciphers in the same period.[18] Reading such authors, as Johannes Trithemius,[19] Giambattista Della Porta,[20] Gustavus Selenus[21], Blaise

[17] For Sukhotin's algorithm that I used to identify the vowels, see Jacques B. M. Guy, "Vowel Identification: an Old (But Good) Algorythm" *Cryptologia* 15/3 (1991): 258-261; and Caxton C. Foster, "A Comparison of Vowel Identification Methods," *Cryptologia* 16/3 (1992): 282-286. A sign of the efficiency of this method is that it can be efficiently used for such esoteric languages as Georgian, Gaelic, and Hungarian. See: George T. Sassoon, "The Application of Sukhotin's Algorythm to Certain Non-English Languages," *Cryptologia* 16/2 (1992): 165-173. However, I do not know of any attempt of applying this method to homophonic systems, that is why I tried myself it manually on several 17th century homophonic cipher systems, and the results were surprisingly good, the more frequent vowels were efficiently identified.

[18] Fort the classic overview, see David Kahn, *The Codebreakers: The Comprehensive History of Secret Communication from Ancient Times to the Internet* (New York: Scribner, 1996), 106-188.

[19] Johannes Trithemius, *Polygraphiae libri sex* (Oppenheim: Haselberg de Aia, 1518), *Steganographia: ars per occultam scripturam* (Frankfurt: Becker, 1606)

[20] Giambattista Della Porta, *De furtivis literarum notis vulgo de ziferis liber quinque* (Naples: Johannes Baptista, 1602), *De occultis literarum notis, seu artis animi sensa occulte aliis significandi*, Starssbourg: Zetzner, 1606).

[21] Gustavus Selenus was the pseudonym of Duke August of Braunschweig (Gustavus being the anagram of his name, Augustus, and Selenus being an allusion to his home town: Lüneburg), name giver of the famous Herzog

Vigenere,[22] and Falconere,[23] one is struck by the great variety of techniques they provide. A considerable portion of these consists of simple monoalphabetic substitutions or transpositions of the letters of the plain text – methods that can be easily cracked by frequency analysis and vowel identification algorithms. Several of these authors provide polyalphabetic ciphers as well, that – as has been mentioned above – does not apply for the Codex of Rohonc, and therefore we do not treat it here.

They also present various ways of hiding the plain message in a longer text (a method that is called steganography, strictly speaking). Of this text only certain letters are supposed to be read (the initial letter of every second word, for example). Being such a popular method in Selenus for example, who devoted long sections to it (differentiating between methods in which every initial counts, methods in which every second, and methods in which all initial count, but the initials of the first and the last words in a sentence are non-significant),[24] it seemed worthwhile to try it on the Codex of Rohonc. Even though the codex does not contain letters but characters, many of them (especially when mirrored) look like letters. My attempts at identifying certain characters as significant, and the rest as non-significant letters, however, remained fruitless, partly because many plain letters of the real alphabet do not seem to appear in the codex (not even in their mirrored form), and partly because this would not account for the strong repetitions of the codex. To be sure, one can never exclude that only one letter or one character per page, or per line counts, or that frequency analysis is to be applied to every 20th character, because only these are significant, and all the rest of the cipher is to be ignored. Nevertheless, this method – being fairly primitive, and yet able to embitter the life of the code breaker – does not seem very likely regarding the strong and repeating structures of the sign strings.

Other methods occurring in these early modern "cryptology monographs" are syllable methods, in which single characters of the cipher text stand for pairs of letters of the plain text, or double characters stand for letters or double characters, or three characters stand for

---

August Bibliothek, who was particularly interested in cipher systems. His basic survey on the topic is entitled *Cryptomenytices et cryptographiae libri IX* (Luneburg, Sternen, 1624). On Duke August, see also Gerhard Strasser, "The Noblest Cryptologist: Duke August the Younger of Brunswick-Luneburg (Gustavus Selenus) and His Cryptological Activities" *Cryptologia* 7, no. 3 (1983): 193-217; idem, "Die kryptographische Sammlung Herzog Augusts: Vom Quellenmaterial für seine Cryptomenytices zu einem Schwerpunkt in seiner Bibliothek" *Wolfenbütteler Beiträge* 5 (1982): 83-121.

[22] Blaise de Vigenère, Traicte des Chiffres (Paris: Abel l'Angelier, 1586).

[23] J. Falconer, *Rules for explaining and decyphering all manner of secret writing, plain and demonstrative with exact methods for understanding intimations by signs, gestures, or speech: also, an account of the secret ways of conveying written messages, discovered by Trithemius, Schottus, Lord Fran. Bacon, Bishop Wilkins, etc.* (London : Printed for Dan. Brown ... and Sam. Manship, 1692).

[24] Selenus, *Cryptomenytices,* Book 3.

three letters, etc.[25] Again, these algorithms do not seem to be applied directly in the system of the script of the Codex of Rohonc, however, looking at the length of the possible "words" in the Codex, and at the many double and triple characters in the cipher, it is highly possible that what we are looking for is a more complex version ultimately based on a syllable method.

Getting thus familiar with the *theory* of ciphers, it might be useful to turn to its actual *practice*. In sharp contrast of the variety of methods provided in the early modern summaries, almost all the methods applied in 16th-17th century ciphering practice belong to one singly category – actually quite neglected in the contemporary monographs – the category of homophonic systems.[26] (Not homophonic ciphers used at the time were very simple methods in which numbers or signs constituted one single cipher alphabet, and these methods can be cracked quicker by brute force than by any systematic method.) The actual homophonic cipher are composed of numbers or special signs falling in one of the following categories: 1) three or four homophones belonging to each letter of the plain alphabet, 2) separate characters to signify double letters, 3) characters to signify syllables, 4) a number of nullities, that is dummy characters without meaning, and finally 5) a set of nomenclators, that is, code words corresponding to the most frequent particles of the given language, and to the key names, geographical and political unities to be mentioned in the given political context. Three of four homophones per letter make up for an alphabet of almost 100 characters, the number of nulls does not exceed usually ten, a dozen of characters are reserved for double letters, syllables often take as many as 100-150 signs, while the dictionary of nomenclators can count more than 300 elements. This was the usual system followed in early modern Italian,[27] Spanish,[28] French,[29] German,[30] and even Hungarian[31] diplomatic correspondence.

---

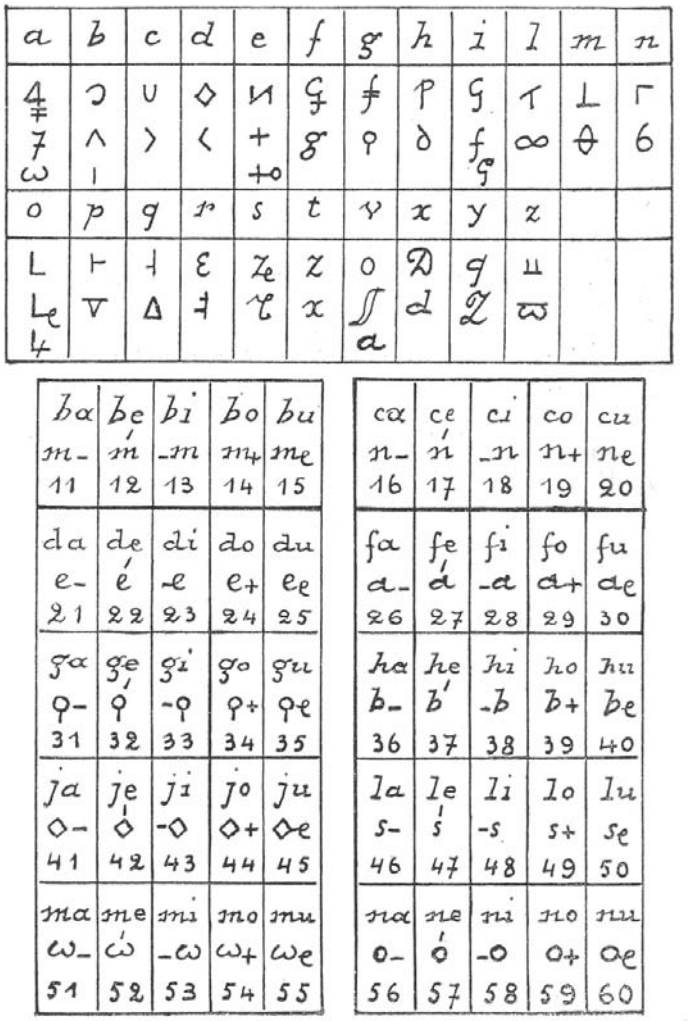
[25] See, for example, Vigenère, Traicte des Chiffres, 238, and Selenus, *Cryptomenytices,* Book 5 Chapter 19.

[26] The discrepancy between the theoretical literature on ciphers and their actual reality was already pointed out by David Kahn, who wrote that the formers "have certain air of unreality about them." Kahn, *Codebreakers*, 156.

[27] Aloys Meister, *Die Geheimschrift im dienste der päpstlichen kurie von ihren anfängen bis zum ende des 16. Jahrhunderts* (Paderborn: Ferdinand Schöningh, 1906); idem, *Die Anfänge der modernen diplomatischen Geheimschrift* (Paderborn: Ferdinand Schöningh, 1902); Liisi Karttunen, *Chiffres diplomatiques des nonces de Pologne vers la fin du XVIe siècle: Extraits des archives des princes Chigi à Rome,* Annales Academiae Scientiarum Fennice (Helsinki: 1911); Luigi Pasini, „Delle scritture in cifra usate dalla Repubblica Veneta," in *Il Regio Archivio Generale di Venezia,* (Venezia: Pietro Naratovich, 1873). 291-328; Gaetano Platania, „La Polonia nelle carte del cardinale Carlo Barberieni Protettore del regno," *Accademie e Biblioteche d'Italia* 56 (n. s. 39) (1988) n. 2. 38-60; Bartolommeo Cecchetti, "Le scritture occulte nella diplomazia veneziana," *Atti del Regio Istituto Veneto* 14 (1868-69): 1186-1211.

[28] J. P. Devos, *Les chiffres de Philippe II (1555-1598) et du Despacho Universal durant le XVIIe siècle*, (Brussels: Académie Royale de Belgique, 1950); Henry Biaudet, *Un chiffre diplomatique du XVIe siècle: Étude sur le cod. Nunz. Polonia 27. A. des archives secretes du Sant-Siège,* Annales Academiae Scientiarum Fennice (Helsinki: 1910); Pierre Speziali, "Aspects de la cryptographie au XVI siècle," *Bibliothèque d'humanisme et Renaissance* 17 (1955): 188-206.

[29] J. P. Devos and H. Seligman, eds. *L'Art de Deschiffrer: Traité de Déchiffrement du XVIIe Siècle de la Secrétairerie d'Etat et de Guerre Espagnole*. Belgium: Université de Louvain, 1967.



The first page of a 16th century homophonic cipher[32]

Even though this method seems to be quite outdated in our post-Enigma period, one has to admit that such homophonic systems using special characters to syllables (and sometimes several characters, that is, homophones, to the same syllable) are not that easy to decipher. Usually only a solid knowledge of the historical background and a correct identification of the language of the plain text let the cryptanalyst succeed. The life of this cryptanalyst is easier of course, if – as it often happens – the author of the text makes

[30] Ludwig von Rockinger: "Über eine bayerische Sammlung von Schlüsseln zu Geheimschriften des sechzehnten Jahrhunderts," *Archivalische Zeitschrift*, 1892: 21-96; Franz Stix, "Die Geheimschriftenschlüssel der Kabinetskanzlei des Kaiser," *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Philologisch-Historische Klasse,* Neue Folge, Fachgruppe II, 1936: 207-226, and 1937: 61-70.

[31] Péter Tusor, "Pázmány bíboros olasz rejtjelkulcsa: C.H. Motmann 'Residente d'Ungheria': A római magyar agenzia történetéhez" (Cardinal Pázmány's Italian Codebook: C. H. Motmann 'Residente d'Ungheria,' On the History of the Hungarian Agenzia in Rome), *Hadtörténelmi közlemények* 116 (2003) 535-581; Zoltán Révay, *Titkosírások. Fejezetek a rejtjelezés történetéből* (Ciphers, Chapters from the History of Cryptology) (Budapest, Zrínyi Katonai Kiadó, 1978.)

[32] Devos, *Les chiffres de Philippe II*, 92.

mistakes, leaves spaces between the words, mixes the cipher with plain text written in the original language, and does not make use of all the possible homophones, but remains satisfied with just one character per letter.

That high level deciphering practice and solid background knowledge is necessary to break this type of cipher becomes obvious if we take a look at the – fairly sporadic – examples of early modern summaries of cryptanalysis.[33] Such summaries, both in the early modern era and in the modern times, remain hidden for most of the contemporaries; professionals who decipher cryptic texts and crack codes, are usually silent about the methods they follow.[34]

Antonio Maria Cospi, secretary of the duke of Toscana, the author of such a handbook explaining the methods of how to break ciphers, honestly limits his competence to simple monoalphabetic ciphers, and is obviously reluctant to deal with homophonic systems, that he and his contemporaries call "chiffres composées." He openly admits that it is almost impossible to solve this letter type,[35] and describes in the followings monoalphabetic ciphers, gives methods to find vowels in syllables, and offers tables that contain the most frequent syllables in French, Latin, and Spanish.

Other authors are fortunately more of an undertaker character, and do not refrain from composed ciphers. François Viète (1540-1603) proposes his "infallible rule" as a method to analyze the triads and dyads of the cipher symbols systematically in order to locate the vowels. This is easier in simple monoalphabetic ciphers, and a bit more complex procedure in homophonic ciphers.[36] Even though Viète starts his description by (correctly) pointing out that Spanish ciphers usually apply separate signs for syllables and a number of code words, his method does not really work for such ciphers unfortunately, unless the very strong formal structure of the cipher text helps the code breaker. We can rightly suppose that Viète, a talented mathematician and an experienced code breaker, was fairly successful solving such ciphers, but the "rules" described in his short text do not equip the reader with truly powerful methods.

---

[33] Let me remind the reader, that the very word cryptanalysis was forged by William Friedman in the 20th century.

[34] Interestingly, historian of cryptology J. P. Devos complains as late as in 1967, that nothing can be known about the cryptographic systems used in 1940-45.

[35] Antonio Maria Cospi, *L'interpretation des chiffres ou reigle pour bien entendre et expliquer facilement toutes sortes de chiffres simples* (Paris: Courbes, 1641), 3: „Or comme il y a deux sortes de chiffres, les uns simples, et les autres composez, laissant à part ces derniers comme presques impossibles à rencontrer et deschiffrer, nous ne parlerons que des premiers quie sont les simples." The text is translated from Cospi's Italian to French by F.I.F.N.P.M. (father Niceron).

[36] Peter Pesic, "François Viète, Father of Modern Cryptanalysis – Two New Manuscripts," *Cryptologia*, 21/1 (1997): 1-29.

The most detailed and didactic handbook of code breaking of that time – to my knowledge – is the anonymous *Art de deschiffrer*,[37] a 136 page handbook written in French in the 17th century. It does not provide "infallible" rules but offers instead a great variety of observations, maxims, rules, and precepts to identify the language of the source, to identify which sign stands for a letter and which for a syllable, to find the nulls, to analyze the frequency of letters and syllables, etc. Mastering these, an attentive reader might successfully attack a contemporary cipher. The author makes the usual differentiation between simple and composed ciphers, treats both at lengths, and gives in both cases detailed case studies. Analyzing French and Spanish ciphers, he can demonstrate how to apply his maxims and precepts.

To be sure, 20th century methods seem to be even more helpful, but as far as homophonic ciphers of historical reality are concerned, the last three hundred years did not add too much to the methods of deciphering. The most useful new methods that a historian might find helpful concern vowel identification. Anyone dealing with 17th century diplomatic ciphers in French and Spanish (but probably also in Italian, and to a smaller degree: German) will find the *Art de deschiffrer* handbook useful, and probably sufficient to break the given cipher. The reason why it does not seem to be satisfactory to solve the Codex of Rohonc is that we do not know the plain language, the text is not that formalized (or more precisely, it is formalized, but we do not know its form), and that we cannot be sure whether it really follows the homophonic and syllable methods detailed above. A *pro* argument is the great number of the signs used in the codex, and that it is not easy to determine this number since new and new signs keep turning up – a characteristic that might be caused by the introduction of new and new nomenclators. Such a method would also account for the relative shortness of the units that seem to be words – if the signs stand for syllables, enciphered words should not be long. In a homophonic cipher, nothing is less curious that the first and the last letter of the inscription INRI on the cross differ (and that on another picture, they differ again). This theory explains furthermore why doubled characters occur so rarely: homophonic ciphers use separate characters for double letters in order to make them invisible.

This hypothesis does not explain, however, that some signs prefer to occur only in the company of some other signs, nor that some sign combinations appear simply too often. Also, I did not manage to find similar sign strings that would differ only in a couple of signs and

[37] Devos and Seligman, *L'Art de Deschiffrer*. See also H. Seligman, "Un traité de déchiffrement du XVIIe siècle," *Revue des Bibliothèques et Archives de Belgique* 6 (1908): 1-19.

that stand for the same words enciphered with slightly different homophones. In contrast, it seems that whenever similar sections are repeated in the text, the signs are precisely the same.

For the moment, the cipher of the Codex of Rohonc resists my attempts based on both historical and modern methods. If anything, it seems to be the closest to a homophonic cipher in which syllables have separate signs, but it might just as well be a writing composed of consonants exclusively, or a system in which sometimes single, sometimes double, and sometimes longer, composite signs mean something – a letter, a consonant, or a syllable.

Meanwhile, we should not forget the possibility either, that the text might consist of codes entirely – not just partially as homophonic ciphers using nomenclators do. As such a system – that can be solved only with huge efforts and some chance – raises the same problems as artificial languages do, we will treat it later.

Shorthand systems

The motives of a scribe writing something in shorthand (speedwriting, stenography[38]) are fairly far from, if not opposite to, the reasons why someone decides encipher a message. And yet, looking at a text written in shorthand and not having the table with the meaning of the signs is strikingly similar to looking at a cipher text but not having the key to it. Deciphering such a text is not always simple, even if its author did not wish to hide anything.[39] Not only the puzzled reader has similar feelings, the secondary literature of the two genres also often coincides,[40] and not rarely the same inventor proposes both a system of shorthand and a cipher. The "father of modern shorthand", John Willis, gives as one of the merits of his stenography, that it is "secret enough to all that are not acquainted with this art",[41] and having presented what he calls "stenographie or compendious writing", he goes on describing the five rules of what he calls "stenographie or secret writing". To these overlaps, we can add a few interesting cases, when certain inventors did not even decide whether they created a shorthand or a cipher – as is the case of several quasi-secret autobiographies, written with one eye at the posterity who will be hopefully able to decipher the not-entirely-private message.[42]

---

[38] Stenography is not to confound with steganography, that is the art of hiding the message.

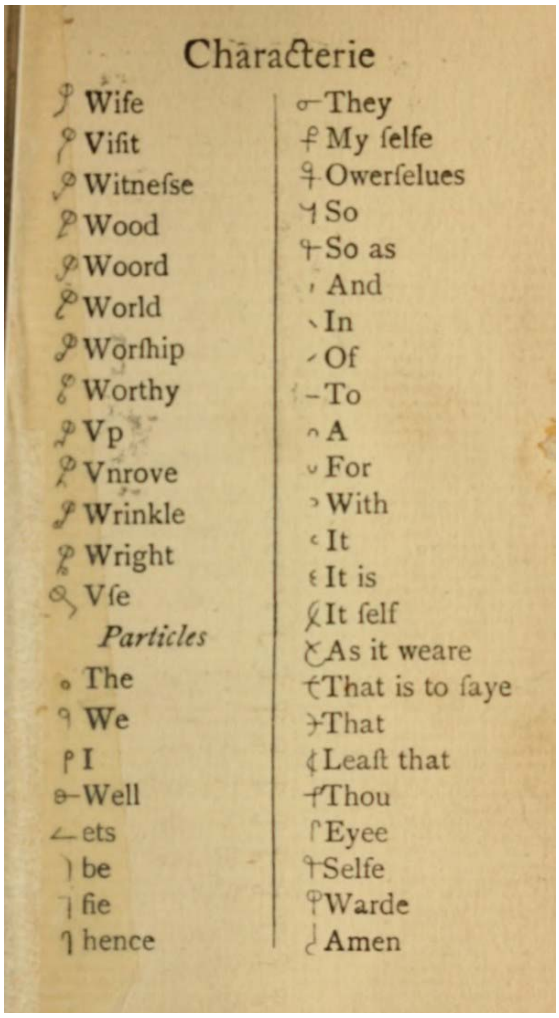
[39] See, for example, James J. Gillogly, „Breaking an Eighteenth Century Shorthand System", *Cryptologia,* 11/2 (1987): 93-98.

[40] Giorgio Costamagna, *Tachiografia notarile e scritture segrete medioevali in Italia*, (Rome: ANAI, 1968), and Falconer, *Rules.*

[41] John Willis, *The Art of Stenographie, teaching by plaine and certaine rules, to the capacitie of the meanest, and for the use of all professions, the way of compendious writing* (London, Cuthbert Burbie, 1602.) This book is extremely rare; I had access to it in the Bibliothèque Nationale de France, Réserve, V 30834. The quoted sentence appears on the second, unnumbered page.

[42] On the autobiography of the Egyptologist, Simeone Levi, see Emanuele Viterbo, "The Ciphered Autobiography of a 19th Century Egyptologist," *Cryptologia,* 22/3 (1998): 231-243. On the secret autobiography

After its renaissance in the 16th century,[43] shorthand enjoyed a fairly complex history, a rich variety of methods flourished. What was common in these systems is that they often used signs not only for the letters of the alphabet, but also for syllables, sometimes omitted vowels and marked only the consonants, applied specific signs for whole words, and frequently used as many as 150-200 characters. Consequently, they face the code breaker with exactly the same variety of problems as early modern ciphers do, and while there are no separate studies on how to break a shorthand system, basically the same cryptanalytic algorithms are to be followed – knowing its plain language does not harm, of course.[44]



A page from Timothie Bright's *Characterie: an arte of shorte, swifte, and secrete writing by character*

---

of the Hungarian writer Géza Gárdonyi, see *Titkosnapló: A titkosírás megfejtői: Gilicze Gábor, Gyürk Ottó.* (Secret diary: solvers of the cipher: Gábor Gilicze, Ottó Gyürk) (ed. Sándor Z. Szalai.) (Budapest: Szépirodalmi, 1974.)

[43] Shorthand, than named as Tironian notes, was very popular until the 11th-12th centuries, using at the end several thousands of signs. In the late Middle Ages they seem to have disappeared, to be reborn in a different form in the late 16th century in the books by John Willis (b.1575) and Timothy Bright (1588).

[44] James Gillogly profits a great deal in the process of code breaking from the fact that the original language of his source is identified, and that several signs had already been deciphered by a previous hand, providing him with „cribs". See Gillogly, „Breaking an Eighteenth Century Shorthand".

(London: T. Windct: 1588)

Is the codex of Rohonc a lengthy application of a shorthand system? Its alleged topic, Christian liturgy does not exclude this possibility. It was quite common to demonstrate the merits of a shorthand system on the Lord's prayer,[45] on various extracts from the Bible, not to mention that in a somewhat later period, a whole gospel got printed(!) edition in shorthand.[46] With this latter publication the author was just about to advertise the system he developed, and thus, he included the list and meaning of the signs in the book. However, without this list, the task of the cryptanalyst would be almost hopeless. It is not easy to recognize which signs are identical, and which are different, in addition, the author has a tendency of combining signs in the same figure, making it very hard for the reader to prepare a complete list of signs. Ostensibly, four kinds of signs are used: for letters, for syllables, for frequent words (such as and, or, we, the, so, to, out, to be, shall be, together), and for often used theme-specific words (Christ, God, heaven, Gospel, world, cross). In this form, this is not exactly the system followed in the Codex of Rohonc, because the latter's sign are not stroke like, they are not particularly easy and quick to draw – a characteristic without which a system of *speedwriting* does not make much sense. In addition, composite signs in the Codex of Rohonc do not seem to be made up from smaller identifiable elements that have their own meaning (as it happens in shorthand), they are just composite in the sense that they are complicated to draw, their constituents, however, cannot be attributed by a specific meaning. Furthermore, as shorthand systems are not really secretive, they do not hide their original language, and include here and there words in plain text, and they often use Arabic numerals openly (though shorthand equivalents of numbers do exist). This is clearly something, that the scribe of the Codex avoids doing, no titles and no numbers appear, nothing that would help the reader identify different parts of the text. Finally, speedwriting – in contrast to the Codex – reads from the left to the right, after all, it is the speed of easy writing and reading that is at stake, making obstacles to the reader belongs more to the domain of secret writing. However, the hypothesis that an author was trying his new system of artificial writing (even without being a quickly

[45] William Fordyce Mavor, *Universal Stenography, or a new compleat system of short writing*, (S. l.: Harrison, s. d.), plate 3 – Lord's prayer, and plate 5: excerpt from the Book of Job. Fort he Lord's prayer in various shorthands: http://en.wikipedia.org/wiki/Shorthand.

[46] *The New Testament of our Lord and Saviour Jesus Christ in Taylor's system of Short Hand as improved by George Odell*, (London: G. Odell, 1843) accessed in Bibliothèque Sainte Geneviève, Paris, Réserve, delta 68 223. Another Bible written in shorthand: *The new testament of our Lord and saviour Jeus Christ, printed in en easy reporting style of phonography by Isaac Pitman*, (London: Frederick Pitman, 1886), accessed in Bibliothèque Sainte Geneviève, Paris, Réserve, delta 68 226.

writeable method) on a sacred text should not be excluded. This, however, leads to the question of artificial language inventions.

Artificial language schemes and codes

Universal language schemes (in other words: philosophical or perfect languages) and pasigraphic projects producing artificial writings had various different, if not contradictory aims in the early modern times. Some of the inventors were still looking at the 'Ursprache' the pristine language in which Adam had named all the things of the world. Some other wished to produce a language in which the true structure of the world could be mirrored. Others did not care about "the nature of the things themselves" and their true relations, they were rather of practical bent, and decided to make a common language and a common writing in which all the nations (and all the religions) would understand each other. Some intended to replace Latin as a medium of communication for scholars, others wanted to base a new language that would unify the grammatical structures of the western European languages, or simply to create a common code language that everyone could read out in his or her own mother tongue.[47] Some authors described never existing languages as part of their book-length utopias,[48] others created a language as part of a larger hoax story.[49] A number of magic and angelic alphabets were made for spiritual communication,[50] while some talented people invented – and keep inventing in our days – languages for intellectual fun, and aesthetic pleasure.[51]

Many important actors of the history of philosophy and science proposed their own artificial language scheme, or at least wrote extensively on the importance of such a project,

[47] Such as Cave Beck, *The Universal Character by which all the Nations in the World may Understand one anothers conceptions, Reading out of one Common Writing their own mother tongues,* (London: Thomas Maxey, 1657), and Athanasius Kircher, *Polygraphia Nova et universalis ex combinatoria arte detecta* (Rome: Varesius, 1663).

[48] For the imaginary languages of Gabriel de Foigny, Denis Veiras, and Simon Tyssot de Patot, see James Knowlson, "The ideal languages of Veiras, Foigny and Tyssot de Patot," *Journal of the History of Ideas*, 1963: 269-278.

[49] George Psalmanazar (1679?-1763), who claimed to be a native of Formosa in 1703, invented a whole language and an alphabet of his alleged home country. See Michael Keevak, *The Pretended Asian, George Psalmanazar's Eighteenth-Century Formosan Hoax*, (Detroit, Wayne State University Press, 2004)

[50] The most famous scripts are John Dee's angelic language and the magic alphabets presented by Cornelius Agrippa, but a great deal of other magic alphabets survived, see Gilles Le Pape, *Les écritures magiques: aux sources du "Registre des 2400 noms" d'anges et d'archanges de Martines de Pasqually* (Milano: Arché, 2006). These alphabets are all simple monoalphabetic substitutions, and thus fairly uninteresting from the point of view of cryptanalysis.

[51] A helpful survey of such contemporary inventions can be found in Sarah L. Higley, *Hildegard of Bingen's Unknown Language: An Edition, Translation, and Discussion* (The New Middle Ages), (New York: Palgrave Macmillan, 2007), see also the following webpages and e-mail lists: CONLANG, Zompist Bulletin Board, Conlanger Bulletin Board.

including Johannes Trithemius, John Wilkins, Athanasius Kircher, René Descartes, Isaac Newton, Gottfried Wilhelm Leibniz, Marin Mersenne, George Dalgarno, Joseph de Maimieux, Francis Lodwick, Cave Beck, or the Hungarian György Kalmár and János Bolyai. Not only the motives of the inventors, but also the looks of their final product differ a lot. Dalgarno's and Newton's schemes are composed of combinations of three or four Latin letters, the languages of Athanasius Kircher and Cave Beck are made of numbers, John Wilkins, Charles Brosses, Joseph de Maimieux, Christian Berger, György Kalmár and others use special characters for their alphabet (pasigraphy), while Johannes Becher applies special characters to signify numbers that ultimately signify Latin words. Some contemporary schemes even use moving characters, that can be "read" only on the screen of a computer.[52] These attempts – while differing in their aims and methods – are strongly interrelated; hence their secondary literature is also common.[53]

As in the case of the shorthand systems, it was not rare that an inventor presented his scheme using the text of the Lord's prayer (John Wilkins, Charles Brosses, Pierre de Bernonville, and even the impostor George Psalmanazar), or that of a biblical text (Francis Lodwick), so again, the topic of the Codex of Rohonc is not alien from such language inventions, even if one has to admit that the texts used to demonstrate the merits of the artificial languages were usually shorter than 450 pages.

How can it be decided whether the Codex of Rohonc belongs to the family of artificial language inventions, or not? One could argue that a perfect language scheme that wishes to show the real nature of the things and the real structure of the world or at least to be a very practical system is composed of very recognizable elements and has an obviously logical structure. In other words, a text written in a perfect language is supposed to be easily decipherable – while the script of the Codex of Rohonc is clearly not. In György Kalmar's system, the sign that stands for 'death' differs in a small addition from that of life, making the system look quite logical, and Francis Lodowick applies similar geometric forms for words having the same grammatical root. However, a close study of the historical examples does not always confirm the idea that perfect languages have transparent structures. Looking at

[52] http://www.omniglot.com/writing/alfakinetix.php; and http://www.omniglot.com/writing/rotor.htm

[53] Monographs on artificial language schemes abound in the last decades, here I mention only a few important ones: Paolo Rossi, Clavis universalis: arti della memoria e logica combinatoria da Lullo a Leibniz (Bologna: il Mulino, 1983) James Knowlson, *Universal Language Schemes in England and France, 1600-1800* (Toronto, University of Toronto Press, 1975) Roberto Pellerey, *Le lingue perfette nel secolo dell'utopia,* (Roma-Bari: Laterza, 1992), Mary M. Slaughter, *Universal languages and scientific taxonomy in the 17th century* (Cambridge: Cambridge University Press, 1982) Umberto Eco, *La ricerca della lingua perfetta nella cultura europea* (Bari : Laterza, 1993), Gerhard Strasser, *Lingua Universalis: Kryptologie und Theorie der Universalsprachen im 16. und 17. Jahrhundert,* (Wolfenbütteler Forschungen, Vol. 38.) (Wiesbaden: Harrassowitz, 1988).

Wilkins' Lord's prayer without the key, for example, is fairly intimidating. True, the same signs stand for the same particles and words, solving it on this basis however would be a considerable success for any code breaker, no inherent "logic" of the language does facilitate the task of decoding. Thus, the simple fact that we stand puzzled looking at the Codex of Rohonc does not in itself exclude that it was written in an artificial language that did not wish to mask the meaning of the text. Nor does the lack of spaces between the words in the codex exclude that. It is true that many artificial languages do not hide word limits, (why would they do so?), however, a number of them do not really care about indicating them more visibly than the Codex of Rohonc does. Nevertheless, if it is a perfect language indeed, where could it possibly belong to in the various sub-branches of such systems?

Several attempts have been made to classify the large literature of artificial language inventions. Distinctions have been drawn between *a priori* and *a posteriori* schemes (based on philosophical considerations and on existing languages, respectively), between imaginary, philosophical, utopist, and international projects, and many further classificatory schemes are also possible. If we try to find a place in this large family of projects for the Codex of Rohonc, it will probably not fall in the category of the quests for the lost Adamic language, nor into that of the more sophisticated philosophical schemes à la Dalgarno or Wilkins that aimed to give a scientific taxonomy of the world.[54] Rather it would belong to the group of the many early 17th century attempts at creating a common writing, where emphasis was not so much on the underlying taxonomy but on practical merits.[55] The primary aim behind these projects was to give a set of written symbols that could be read off in the mother tongue of the reader, hence their denomination: "universal character" or "escriture universelle". That such a system would rely on a very large number of signs (each belonging to a radical word) was commonly recognized – among others by Francis Bacon and John Wilkins.[56] Many such projects were heralded but the details have been withheld and the whole project remained ultimately in secrecy (those of des Vallées and Jean le Maire, for example), and only a few were worked out (Francis Lodwick's *Common Writing* being the most famous).

If this conclusion is accepted, and the Codex of Rohonc is to be viewed indeed as an application of an artificial language scheme, then its characters or combination of characters stand probably for whole words, notions and frequent conjunctive particles, and they do not refer to the letters of any ordinary language. This means, in turn, that it should not be seen as

[54] Slaughter, *Universal languages and scientific taxonomy.*
[55] Knowlson, *Universal Language Schemes*, chapter 2.
[56] Ibid. 53-56.

a cipher, in which the individual letters of the plain text are treated as units, but as a code, in which the words or phrases of the plain text are treated as units. Therefore, solving it does not require methods of cipher analysis – including frequency analysis, vowel identification algorithm, word pattern analysis, and other methods – but rather of code breaking.

Trying to apply common methods of code breaking,[57] however, I cannot report of any further success beyond the already identified character combinations (that stand for Christ, Pilate, and Jerusalem). This is not particularly surprising. William Friedman, the famous code breaker of the 20th century, names as the first of the four basic operations in cryptanalysis: the determination of the language employed.[58] The anonymous author of *Enemy Codes and their Solution* also feels necessary to point out on the first page of his work, that "The would be solver must possess a thorough knowledge of the language employed, not only from the point of view of vocabulary but also from that of a knowledge of all the peculiarities of its grammar, syntax and idiom."[59] This is not a real problem in normal wartime when the language of the enemy is usually known (the case of the Navajo code talkers being a famous exception). In the case of the Codex of Rohonc, however, let our task be cipher analysis or code breaking, this step is highly problematic, as most of European languages cannot be excluded.

One can argue of course that artificial language inventions in the 16th – 17th century are not known in too many languages, a consideration that narrows down the possible plain languages to Latin, German, French, and Italian, but one cannot exclude either that unknown projects were carried out in further languages, or that the codex was born in a slightly later period, when artificial inventions were made in many other languages. Another counter argument that seemingly avoid the linguistic difficulty might rely on the explicit intent of many common writing projects to become "supra-national". Francis Bacon, John Webster and

---

[57] "Enemy Codes and their Solution," *Cryptologia* 19/2 (1995): 166-195; L.F. Safford, "The Functions and Duties of the Cryptography Section, Naval communications," *Cryptologia* 16/3 (1992): 265-281; *Basic cryptanalysis*, (Field manual 34-40-2) (Washington, DC, 1990), chapters 14-15.

[58] William F. Friedman, *Military Cryptanalysis*, (Laguna Hills: Aegean Park Press, 1980), 1. vol., 7. On the following page, Friedman has an interesting footnote: "The writer has seen in print statements that "during the World War . . . decoded messages in Japanese and Russian without knowing a word of either language." The extent to which such statements are exaggerated will soon become obvious to the student. Of course, there are occasional instances in which a mere clerk with quite limited experience may be able to "solve" a message in an extremely simple system in a language of which he has no knowledge at all; but such a "solution" calls for nothing more arduous than the ability to recognize pronounceable combinations of vowels and consonants–an ability that hardly deserves to be rated as "cryptanalytic" in any real sense. To say that it is possible to solve a cryptogram in a foreign language "without knowing a word of that language" is not quite the same as to say that it is possible to do so with only a slight knowledge of the language; and it may be stated without cavil that the better the cryptanalyst's knowledge of the language, the greater are the chances for his success and, in any case, the easier is his work."

[59] "Enemy Codes and their Solution," 168.

John Wilkins argue that real characters should not express letters or words, but things and notions, universal ideas that appear in all languages in different form.[60] Such a common writing might become universally intelligible for men speaking different languages. The intentions of the designers and reality are two different things, however, and we may plausibly suppose that every artificial language scheme bears the seal of the languages its constructor spoke. All in all, if no language stands behind a code system, it will be even more dubious to solve.

Another practical reason why code breaking is not particularly successful in the Codex of Rohonc is that such methods usually make use of the stereotypical features of the type of text to be broken. In military cases for example, code breakers start breaking the dates, numbers, the first and the last groups in the message, and other parts where one can expect specific information to turn up. Now, the content of the Codex of Rohonc is also stereotypied, and by every means has parts where one can expect specific information to turn up (words, like God, Amen, Christ, Barrabas, Peter, etc), just we do not know where these words are located exactly, and we do not know the exact genre of the encoded text. So while the content of the Codex is probably less vital than that of a World War II message, and it was encoded by a considerably older and simpler method, we do not have such a stable knowledge on the structure of the genre as the military code breakers have on their sources. Since we do not have so good analogues, our code breaking is after all less fortunate.

One could add of course to the two above reasons a considerably simpler one that would explain the lack of success just as well: the problem might simply lay in the fact that no professional code breaker has ever tried to solve the system of the Codex of Rohonc, and it is no wonder that a few amateur code breakers, including the author of the present article, have not yet managed to find its solution.

Instead of solution

Quite similar to the case of the Voynich manuscript, it is easier to determine what the Codex of Rohonc is not, than to say with certainty what it is.

As I have argued thus far, it is not very likely that the Codex of Rohonc is a pseudo-historical source that pretends to be an old Hungarian source, because – with its too many characters and the lack of spaces between the words – it simply does not seem to be a book written in a natural language. Neither does it seem to be a hoax that aims to be sold to a

[60] Knowlson, *Universal Language Schemes,* 16 and 25. For Wilkins, see his *Mercury or the Secret and Swift Messenger,* (London: Baldwin, 1694), 109.

deceived book collector, because – as a book – it does not look very precious, nor particularly mysterious besides that fact that it is written in unknown characters. Nor does it seem to be the product of a crazy mind, first because it actually seems to have been written by several hands, so at best, two crazy minds should be assumed, and second, because the grammatical rules according to the text is structured are very consequent, and thus it was definitely not improvised by some enthusiastic persons, its system was rather consciously designed.

If, however, it has a consciously designed system, what that system might look like? Because of the very strong repetitions of the text (in other words, because of its low entropy), a polyalphabetic encryption is quickly ruled out. It is more likely that the same characters of the cipher text refer always to the same entities (letters, words, or ideas) of the plain text. If this is so, however, three options remain: the Codex contains a monoalphabetic cipher, possibly with homophones, nulls and nomenclators, or it is written in shorthand, or it is encoded in an artificial language.

The problem is that equally strong counterarguments might be raised against each of these three suppositions. In contrast to most of the survived shorthands, it is not composed of sufficiently simple elements, drawing its characters takes time. It would be a fairly slow speed writing.

Homophonic systems – composed of letters or syllables that tried to confuse the code-breakers with the introduction of nulls and nomenclators – flourished in great variety in the 16th-17th centuries. However, more in diplomatic sources than in religious books. Such ciphers were used to make short dispatches secret, never to encipher whole volumes.

Artificial language schemes and common writing projects flourished in even greater variety in the early modern era, and these projects were often displayed in religious texts, but again, we do not have any analogy for choosing such an extensive source for that purpose (a source that cannot even be correctly identified), and the secretive attitude of the encoder (who hides the spaces, and does not use any recognizable character) is not typical for perfect languages.

The only positive conclusion, I could draw, is that whatever system the codex uses, our task is more that of a code breaker than that of a cipher analyst: groups of characters that stand for ideas and words should be broken first, before we find out whether there are at all characters that stand for letters in any language.

We are in a ironic situation: whatever the system of the Codex of Rohonc is, it is fairly outdated, and yet, we cannot solve it. At latest it was designed in the beginning of the 19th century, and codes and ciphers that are so old can be usually broken today. The system is

outdated in the sense that knowing its plain language and historical context a contemporary or a modern historian could probably decipher it with some moderate difficulty. But we cannot solve it, because without its historical context, linguistic background, identification of word limits, etc. a whole office of code breakers could work on it for weeks without being able to decipher it. To put it differently: in theory, the system must be solvable, in practice however, it is not.

When the famous and successful code breaker, William Friedman was asked whether he believes that the infamous and unsolved Beale Ciphers, that are believed to unveil the location of hidden treasure, are real ciphers or not, he answered: "On Mondays, Wednesdays and Fridays, I think it is real, on Tuesdays, Thursdays and Saturdays I think it is a hoax."[61] Quite similar are the enthusiasm and disappointment of János Jerney the first scholar who turned considerable energy around 1842 to understand the system of Codex of Rohonc: "As far as I am concerned, I am touching it eagerly, thumbing it ever and again, I analyzed it several times, made comparisons, sometimes fantasizing about useful results, and then, discouraged by its readability, I always put aside this extraordinary curiosity."[62] Almost 170 years have passed, but the words Jerney gave about his own mixed feelings seem to provide a precise description of what anyone might feel after a few hour work on the Codex of Rohonc.

Is this to say that nothing has happened in the almost 170 years that passed? True, we have not solved the mystery of the characters of the codex. On the other hand, it is useful to clarify what can be reconstructed on a scholarly basis about a source that is actually used to heal the wounds of several nations' self-esteem. As we have seen, the codex of Rohonc is (ab)used as an evidence to reconstruct both Hungarian and Romanian proto-histories, and these attempts have been strongly criticized by other (respectively) Hungarian and Romanian scholars. Even though we are far from being able to solve the mystery of the codex of Rohonc, we can claim with confidence what it is not, and how it cannot be solved.

[61] Quoted in R. Clark, *The Man Who Broke Purple,* (Boston, MA: Little, Brown and Co., 1977), 126.

[62] János Jerney, *Némi világosítások az ismeretlen jellemű rohonczi írott könyvre* (Some considerations on the written Rohonc book of unknown character), Tudománytár, 8, 1844. 15/ 1, 25–36, esp. 26. "... Mi engem' illet, mohón nyúlék hozzá, ismét és ismét megforgatva, hív hasonmásolatokat is véve, többször vizsgáltam, egybehasonlítottam, már hasznos eredményeket képzelve fejtegettem, majd elcsüggedve olvashatása fölött, mindannyiszor félre tettem a' rendkívüli ritkaságot."