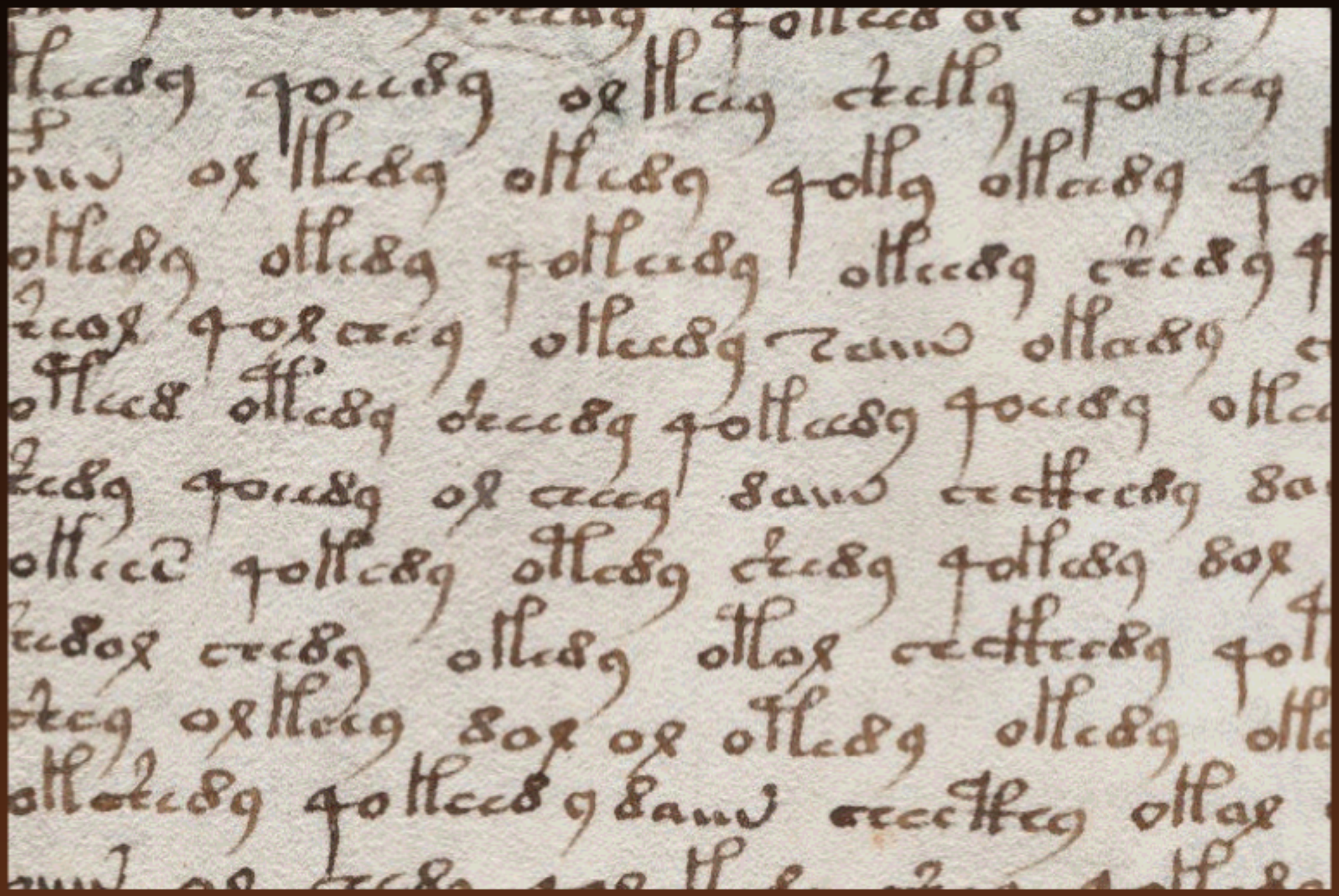
The Voynich Manuscript

– stars, constellations and calendars –

Budapest, 2014 February 4.

The manuscript



Sources – where can we find some more info?

- The manuscript itself: Yale University, Rare Book and Manuscript Library
- Interactive downloading:

  `http://beinecke.library.yale.edu/digitallibrary/voynich.html`

- Lots of material: 206 individual images, one image is 10MB on average (4 – 5 – 6 kpix resolution per side, high quality + lossy compression in JPEG format), all in all: 2 gigabytes
- Get started – sources on the web:
  - René Zandbergen: `http://voynich.nu`
  - Wikipedia articles (and following links)
- Transcription (see later): 260k: character/letter/symbol, 37k: word

The (reversed) history of the manuscript

- 1969: it is moved to its current location (Yale University),
- 1931: widow of Voynich forwards it to one of her friends (A. Nill), who passes it in '61 to an antiquarian (H. P. Kraus);
- 1912: in Villa Mondragone, Wilfrid M. Voynich purchased all of the 30 manuscripts what were on sale, including this one;
- around 1866: the manuscript goes from Collegium Romanumban to Villa Mondragone, which is the private library of Petrus Beckx (a senior Jesuit);
- After the dead of Kirchner, the manuscript was thought to be in Collegium Romanumban for approximately 200 years;
- After the death of Baresch, it gets to Kircher, using the contact person Jan Marek Marci. He mentioned first, that it could be owned by Rudolph II (known by him as a work by R. Bacon).
- First known owner: Georg Baresch (alchemist, 17th century), the Jesuit Athanasius Kircher was also interested in the manuscript but he was unable to retrieve it;
- It is supposed to be written between 1404 and 1438 (based on C14 estimations), a more conservative estimate: in the interval 1350 – 1500. Author(s), language, alphabet, purpose, etc. is unknown.

How does the manuscript looks like?

- Currently: it is good to have some sort of page numbering or something equivalent in order to make references.
- Formerly used units in manuscripts and in typography:
  - *folio*, front page: *recto* back page: *verso*;
  - *cruda* (paper quire);
  - the folios and crudas are enumerated;
- The Voynich manuscripts:
  - 20 quires;
  - 116 folios, including foldouts;
  - 206+ „pages” (depending on if we count foldouts to one ore more page)

  Identification: „f<folio>[r|v](1,2,...)”, e.g. f1r, f68v2, ...
- „Chapters”: based on figures and typography (and/or quires, however, the correlation is not so strict, there are missing folios and there are overlaps as well).

The chapters – I. Herbal

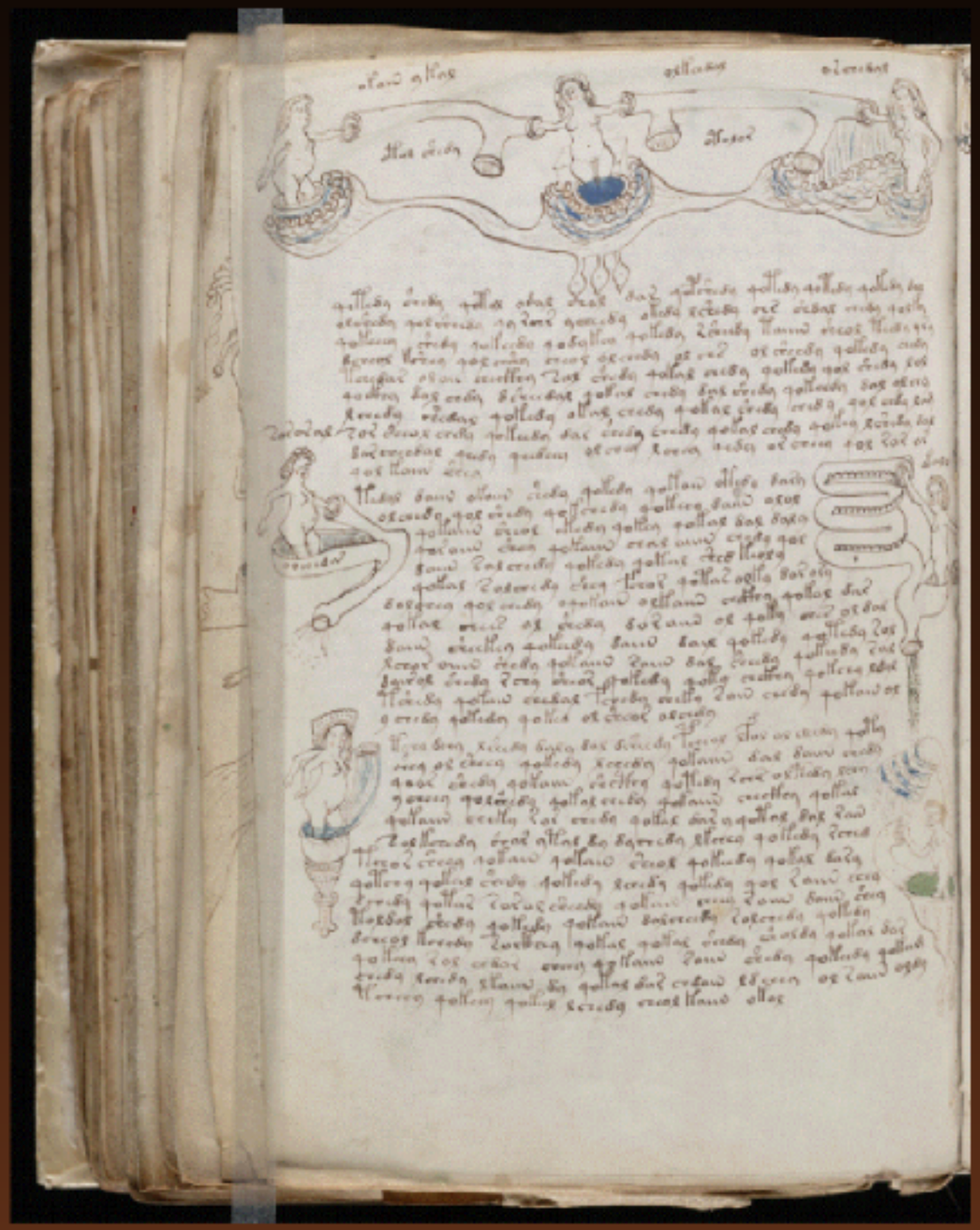
f44v

f45r

The chapters – II. Astronomy

f67r1 *f67r2*

The chapters – III. Biology or Anatomy



f77v

f78r

The chapters – IV. Cosmology

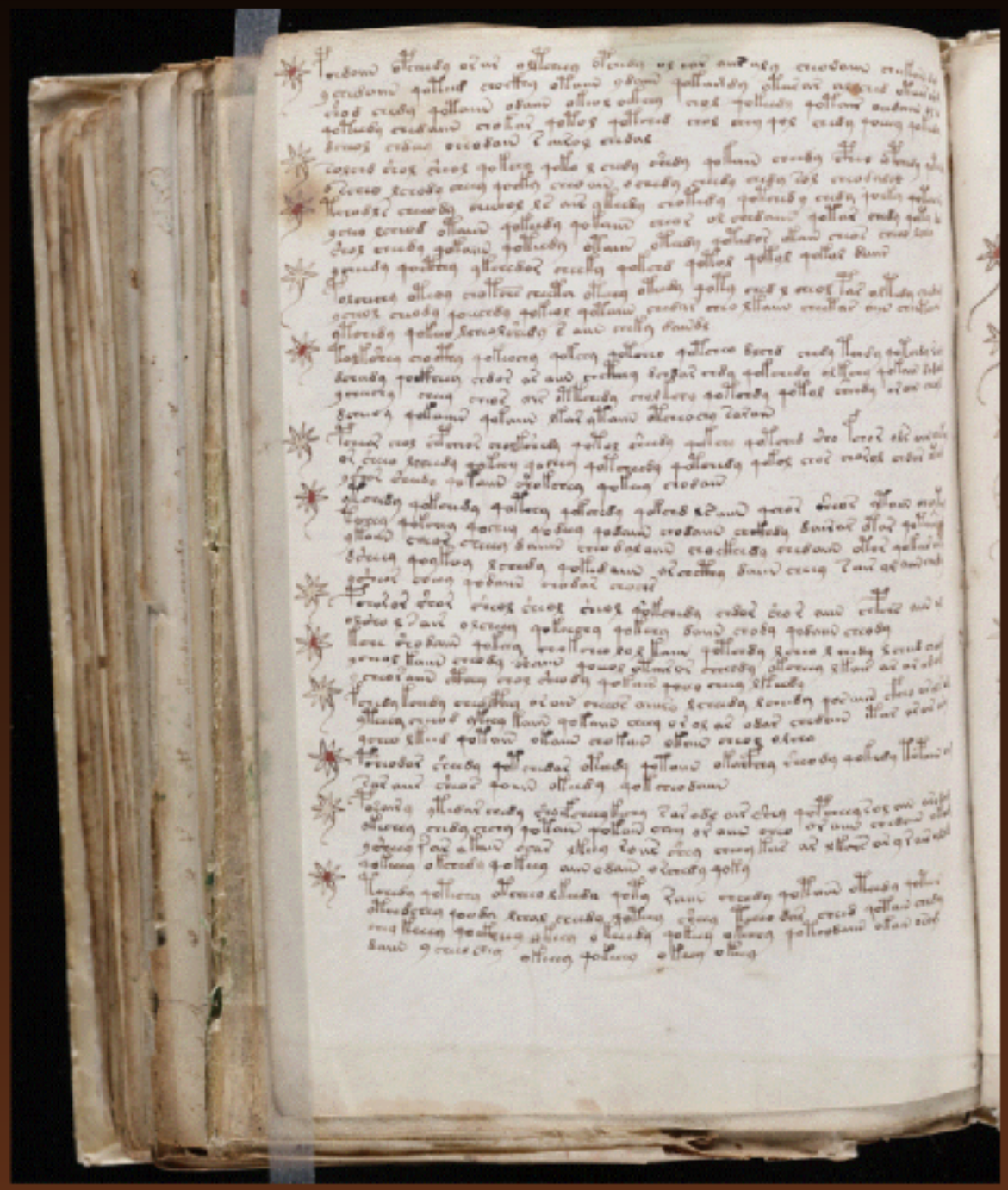
f86v1 . . . v6

The chapters – V. Pharmaceutical

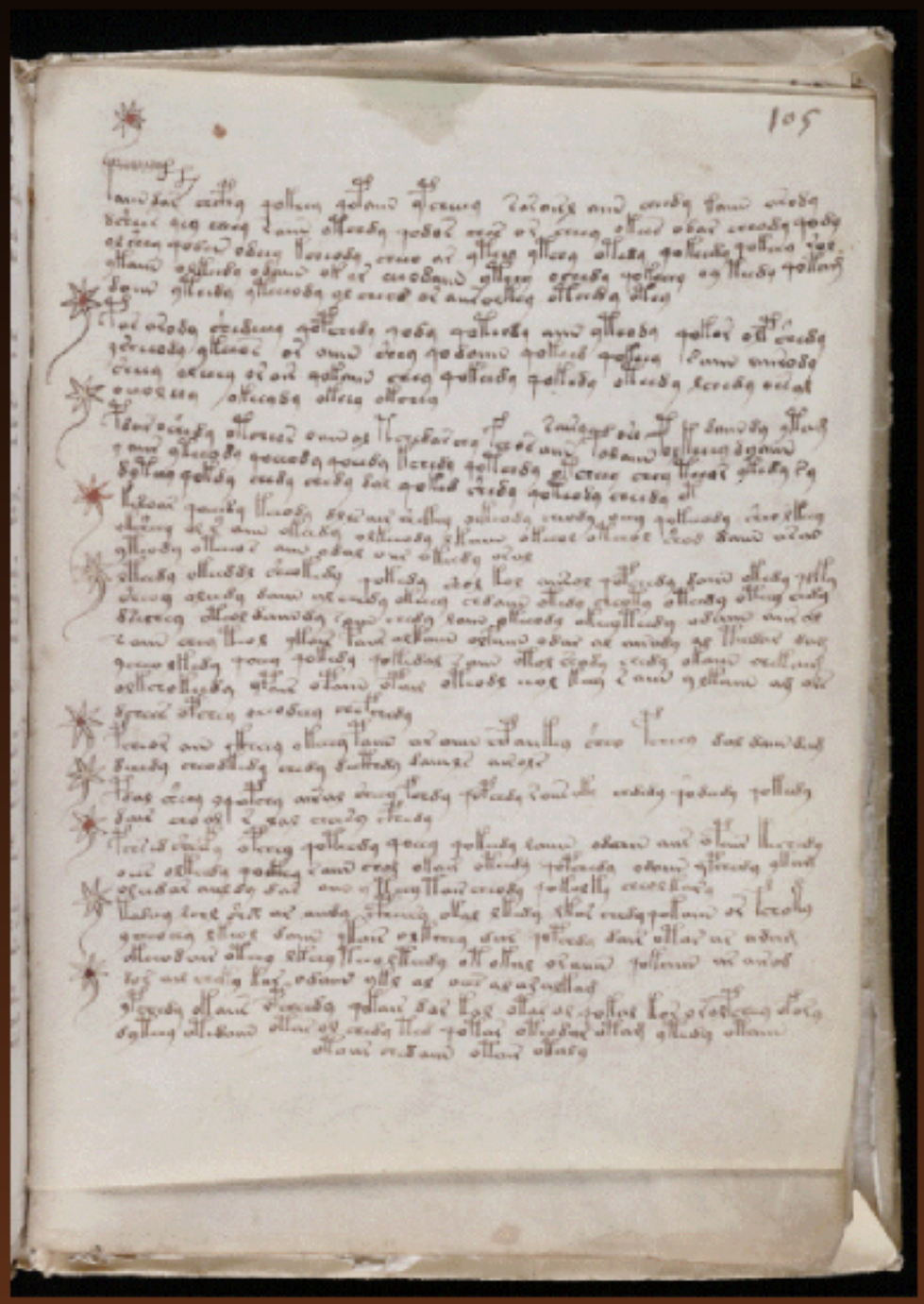
f99v

f100r

The chapters – VI. Recipes



f105v



f105r

Transcription

- „Some sort of a" language ⇒ analyzable text: it would be good to have some sort of transcription...
- Russian, Hebrew, Arabic, Greek: very familiar examples;
- the transcription based on the targeted language (environment):

Soyuz – Soyuz – Szojuz

ישראל – Israel – Izrael

- what makes it hard: units, graphemes, digraphs, ligatures, unreadable text, ...

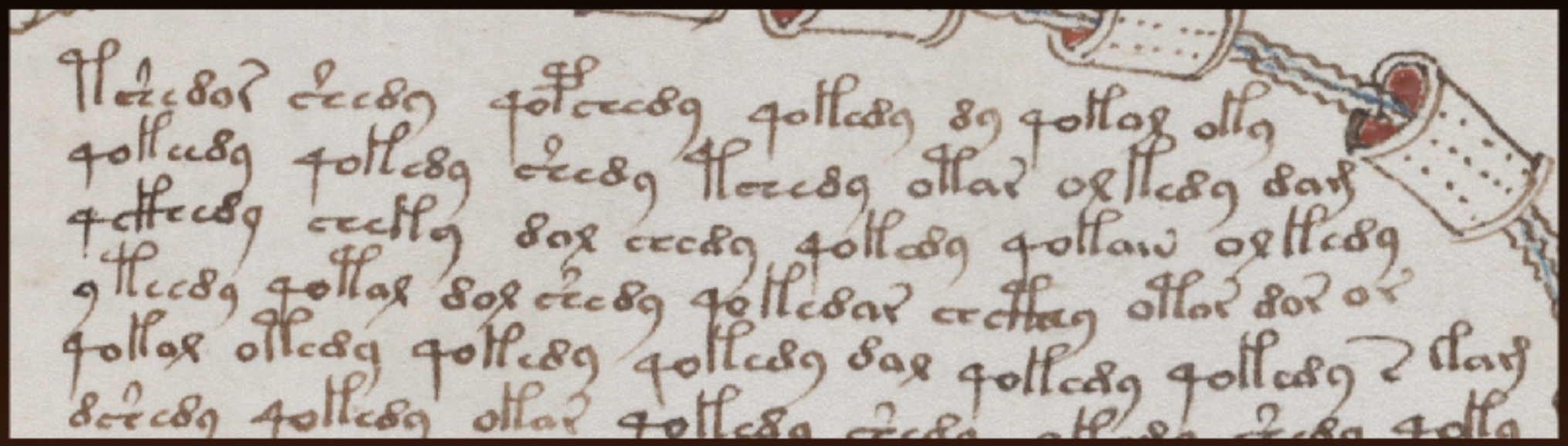
Transcription – the EVA

European Voynich Alphabet (R. Zandbergen; G. Landini, 1998):

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	x	y	z
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	X	Y	Z

- each character (glyph, letter, grapheme, . . .) is transcripted to the most look-alike Latin (or Greek) letter;
- vowels are included: it is good that the transcription is „readable" at some sense.
- one of the most widely known transcription: by Takeshi Takahashi.
- punctuation: almost nothing.
- important question: ambiguity – there are ligatures as well.

Transcription – some examples



tshedor shedy qopchedy qokedy dy qokoy oky
qokeedy qokedy shedy tchedy otar olkedy daim
qckhedy cheky dol chedy qokedy qokaiin olkedy
yteedy qotal dol shedy qokedar chcthey otor dor or
qokol otedy qokedy qokedy dal qkedy qokedy s kam
dshedy qokedy okar qokedy shedy . . .

f78r

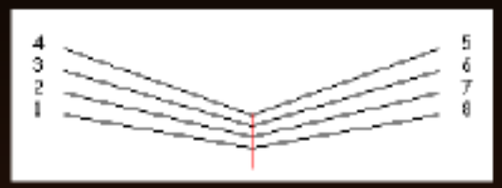
what is typical: length and distribution of the words, repetition of the words, frequency of various letters

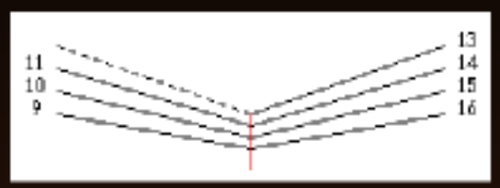
The letters

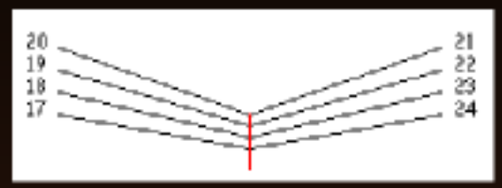
I cdn'uolt blveiee taht I cluod aulaclty uesdnatnrd waht I was rdanieg: the phaonmneel pweor of the hmuan mnid. Aoccdrnig to a rseearch taem at Cmabrigde Uinervtisy, it deosn't mttaer in waht oredr the ltteers in a wrod are, the olny iprmoatnt tihng is taht the frist and lsat ltteer be in the rghit pclae. The rset can be a taotl mses and you can sitll raed it wouthit a porbelm. Tihs is bcuseae the huamn mnid deos not raed ervey lteter by istlef, but the wrod as a wlohe. Scuh a cdonition is arppoiatrely cllaed Typoglycemia.

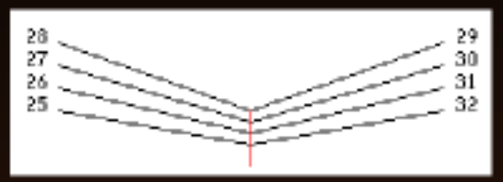
Amzanig huh? Yaeh . . .

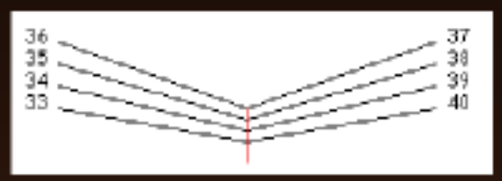
The structure of the manuscript

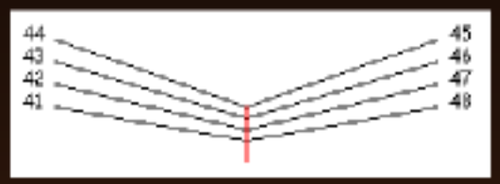


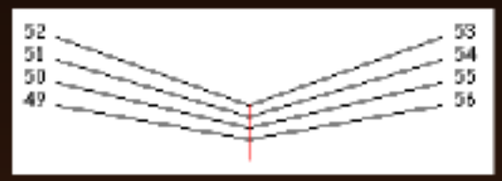


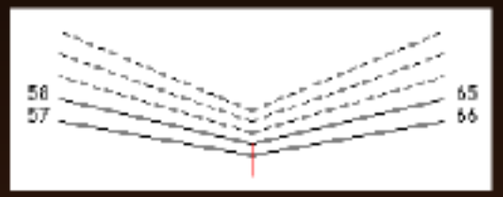


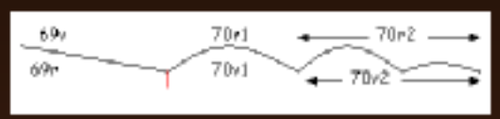


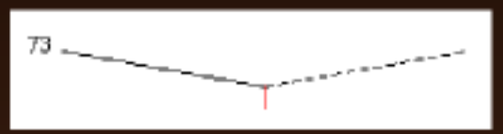


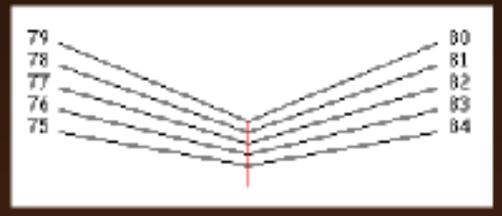


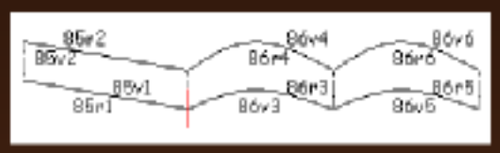


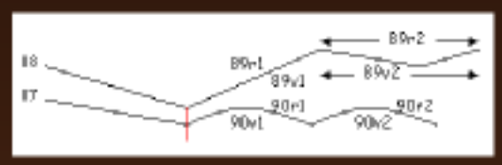


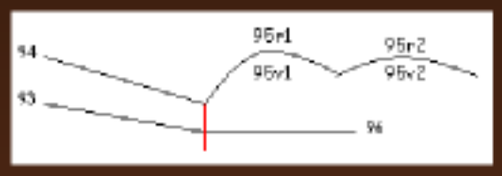


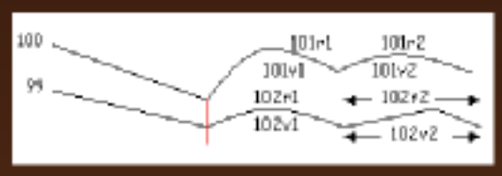


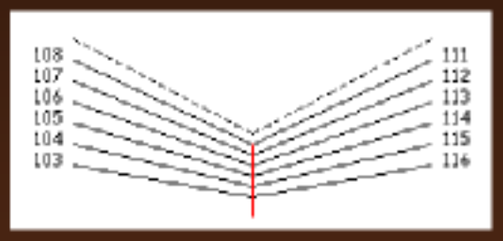












(see also: `http://voynich.nu/layout.html`)

The astronomical chapter

- Pages: f67r1, f67r2, f67v1, f67v2, f68r1, f68r2, f68r3, f68v1, f68v2, f68v3, f69r, f69v1, f69v2, f69v3, f70r, f70v, f71r, f71v1, f71v2, f71v3, f71v4, f72r1, f72r2, f72v, f73r, f73v
- In total: 8 folios, 4 quires (due to the lot of foldouts), 26 pages.
- In a page, following a short text (a kind of „abstract"), there are complex, symmetric figures that can be rotated and there are many words resembling to labels:
  - 10 pages: unambiguously astronomical context;
  - 4 pages more or less astronomy (more resembles to Cosmology);
  - 10+2 pages: calendar (10 months, two are splitted to two parts);
  - these are: zodiac: Pisces . . . Sagittarius (Cap and Aqu are missing)
- in some places, there are additional handwritings (for instance, the Latin names of the months, easy to recognize).
- symbolics: Sun and Moon, stars, text written using red ink (nowhere else); typography resembles to galaxies, naked women having bath (see also: the chapter related to biology, anatomy); one-word labels.

Symmetries

f67r1	12
f67r2	12
f67v1	4
f67v2	19
f68r3	4+4
f68v1	4+4
f68v2	4
f68v3	8
f69r	22
f69v1	14+14
f69v2	8
f69v3	8

Particularly interesting: f67r1 and f67r2.

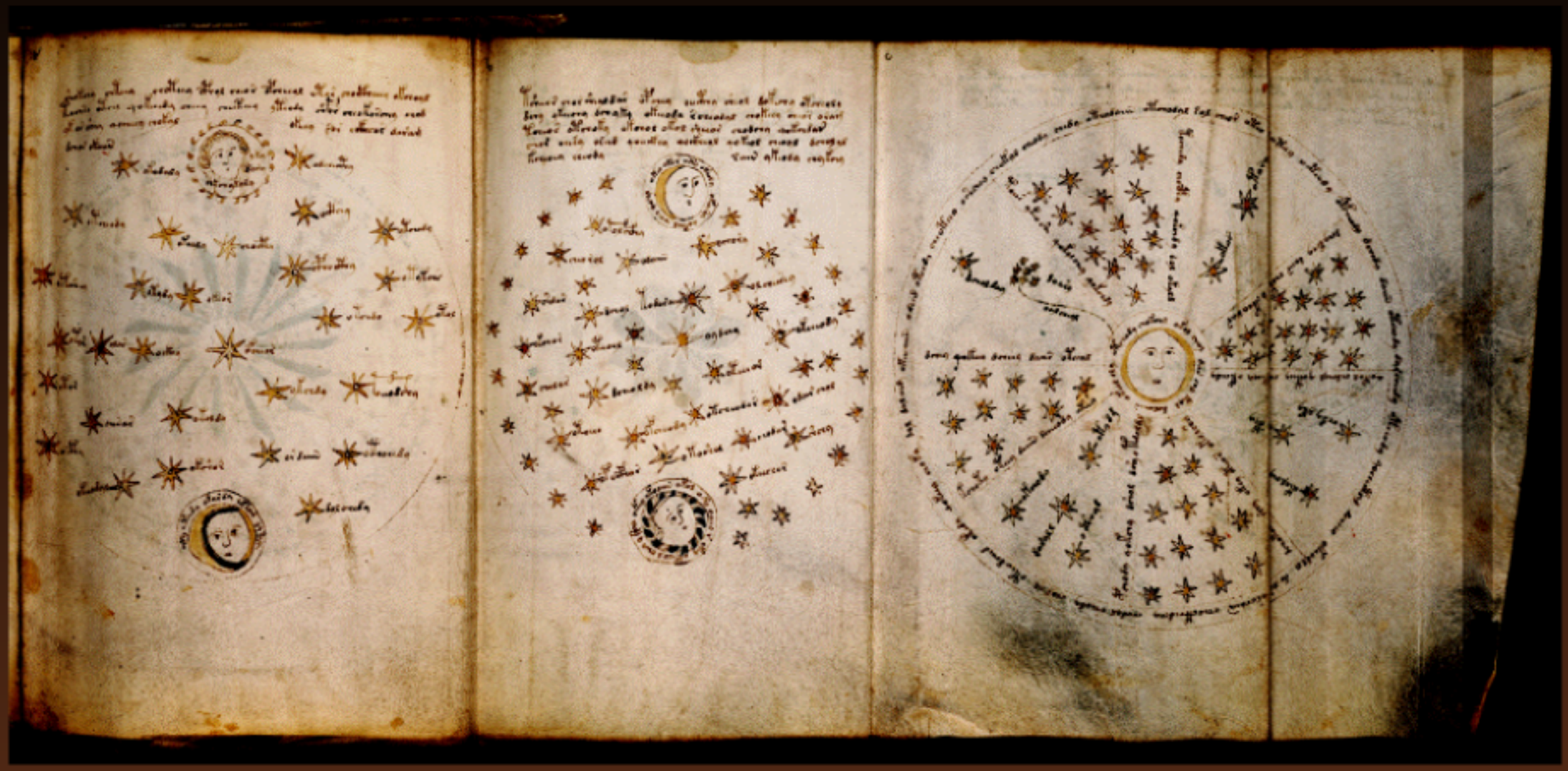
Folio 67 recto

f67r1 f67r2

Properties of the astronomy chapter

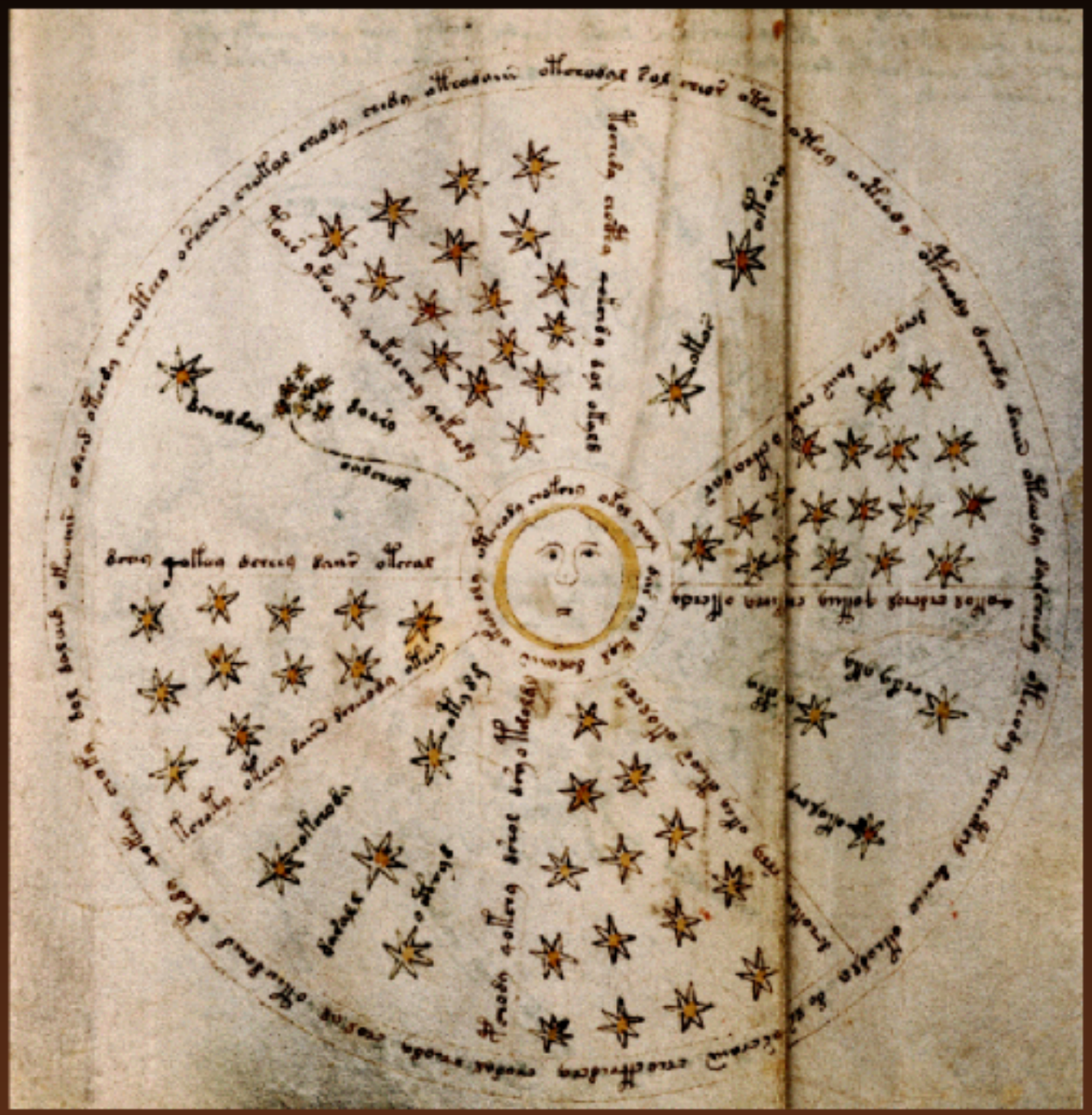
- What is an advantage can also be a disadvantage:
  - lot of figures $\Leftrightarrow$ not so much text;
  - 13k characters in total: how can reliable statistics be made?
- The structure of the text:
  - there are „abstracts" or „captions" (see e.g. sky maps: f68r1, f68r2)
  - labels having the same symmetry group as the figs (even on the first: f67r1)
  - circular or radial text with a dozen of words (see e.g. f68r3)
  - some emphasized words (see e.g. f68v1, f69v2) or letter (see e.g. f69r)

Sky maps



f68r1, f68r2, f68r3

Sky maps – details

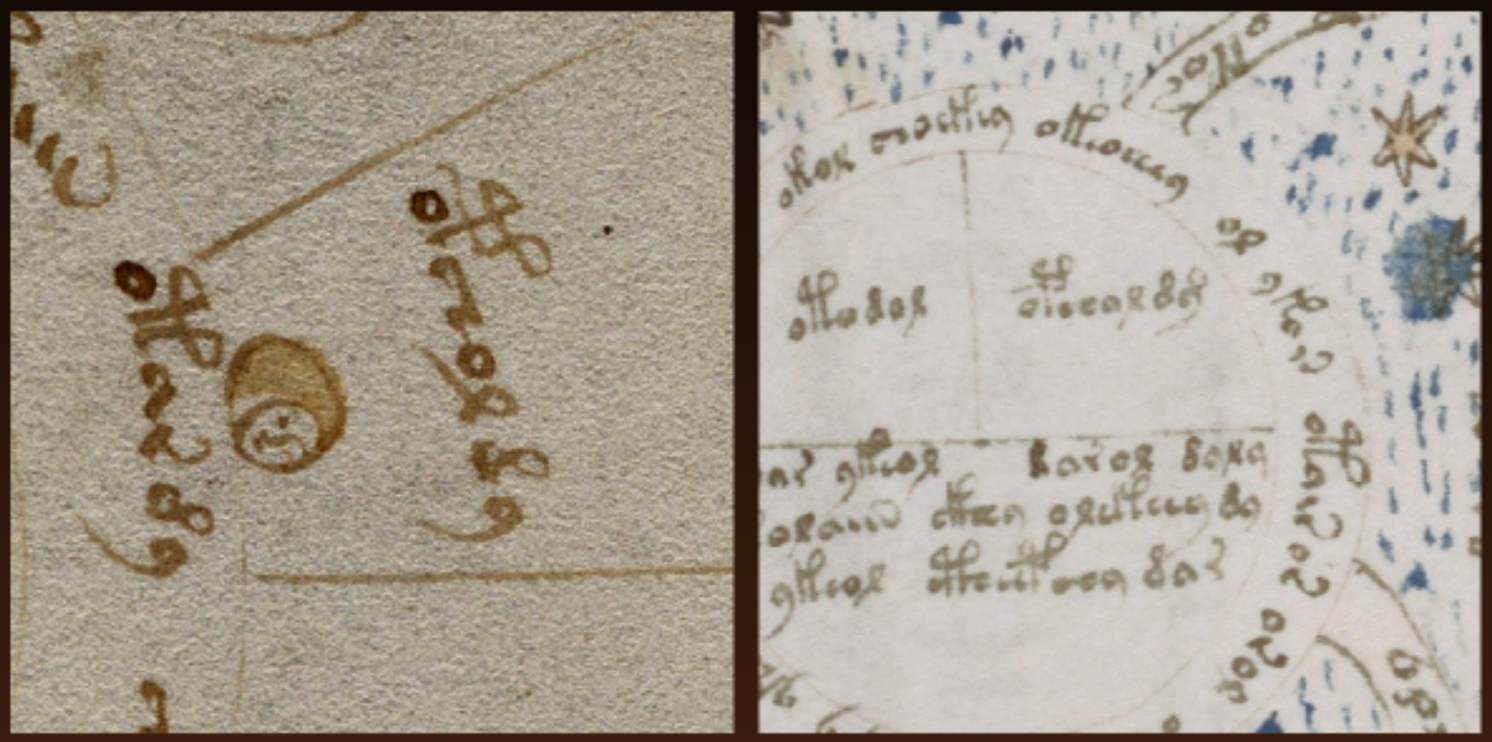


f68r3

Pleiades (Seven sisters) – doary

http://apod.nasa.gov/apod/ap120406.html

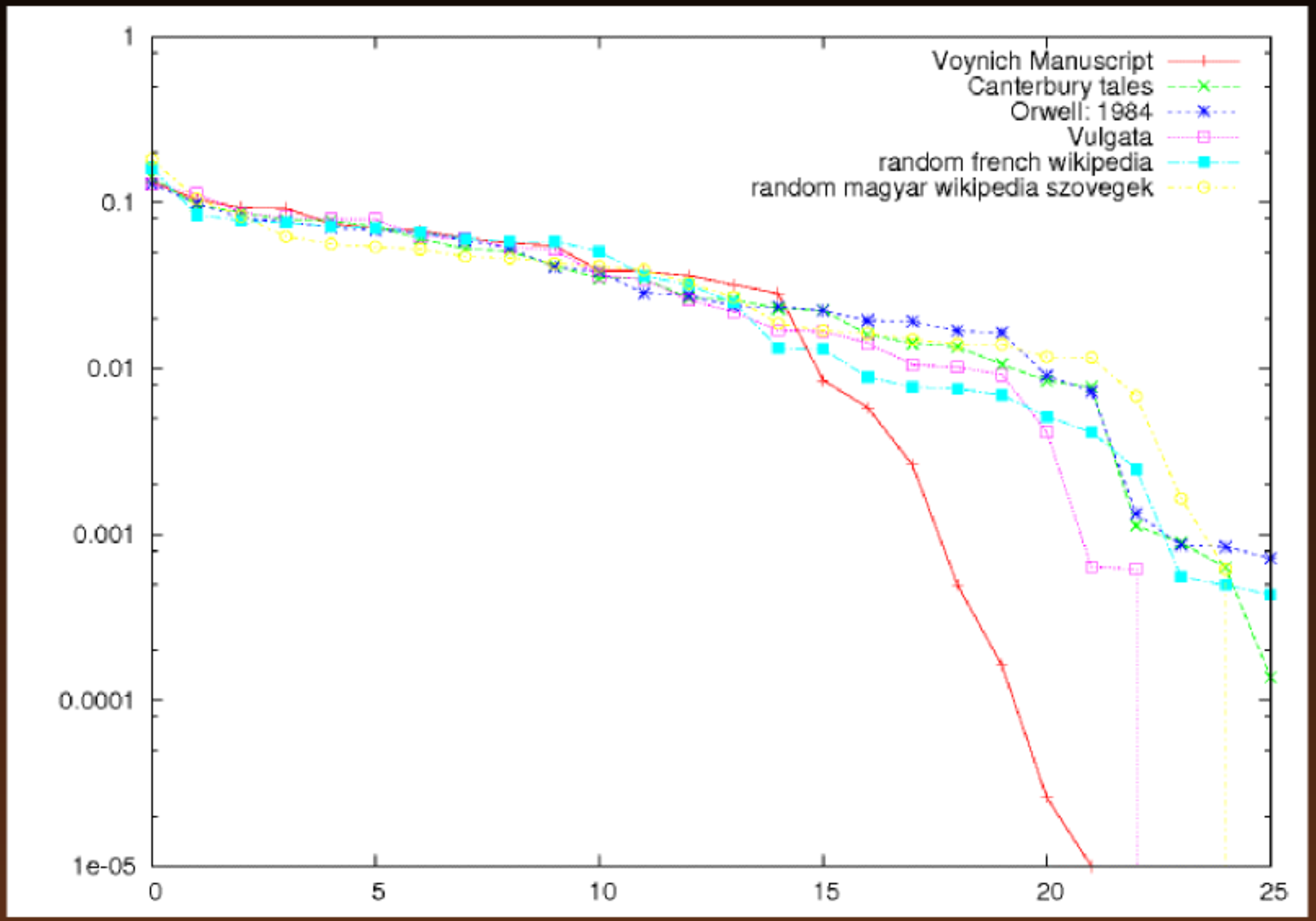
Some prominent matches



- The word: ‘opcholdy‘, it is a) one of the „planets” in f67r2 (in 7 of the 12 sectors there is a single, emphasized word) and b) it can be found at the very center of f68v1 (one of the 4 sectors).
- it does not occur anywhere else, however, the string ‘choldy‘ can be found 17 times at end of words, and there are words like ‘fcholdy‘, ‘ocholdy‘ and ‘yfcholdy‘.

Without the claim of the completeness: there are some more similar words, e.g. ‘otodal‘and some shorter ones (like ‘okal‘)

The letters – Zipf's Law

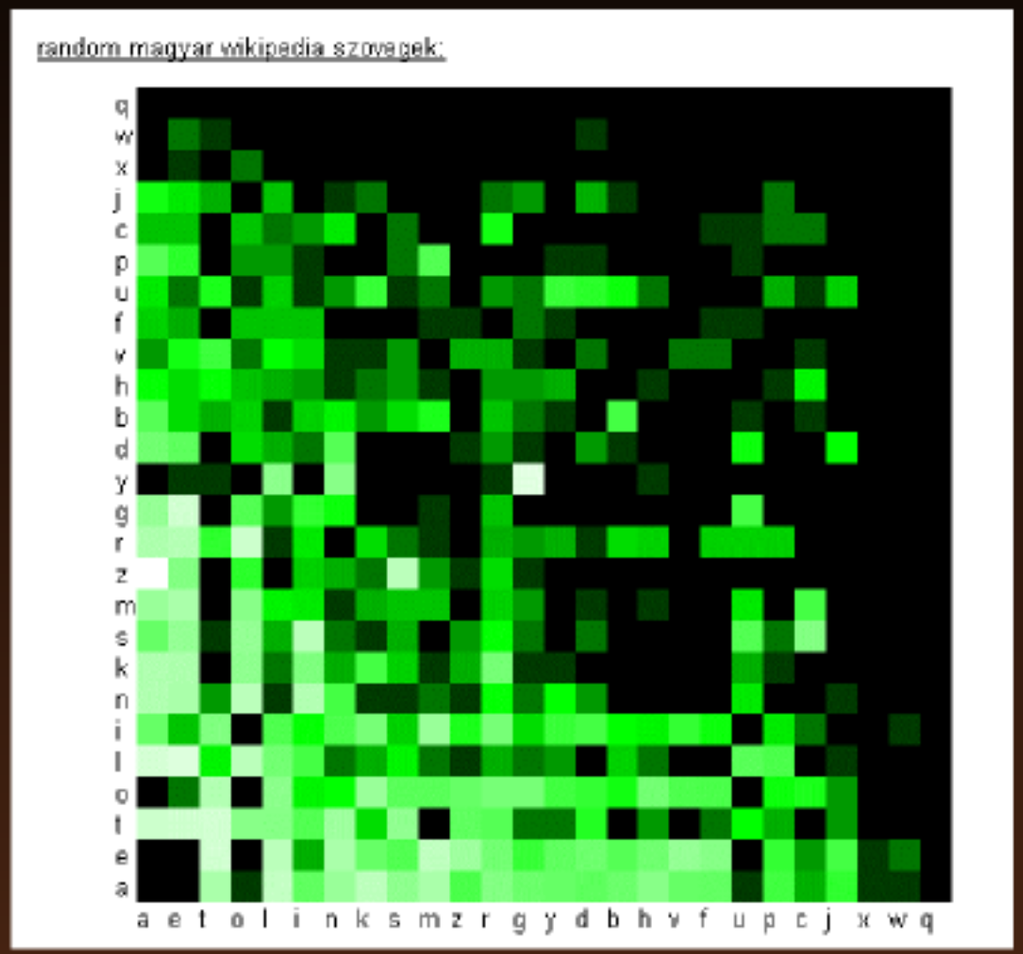


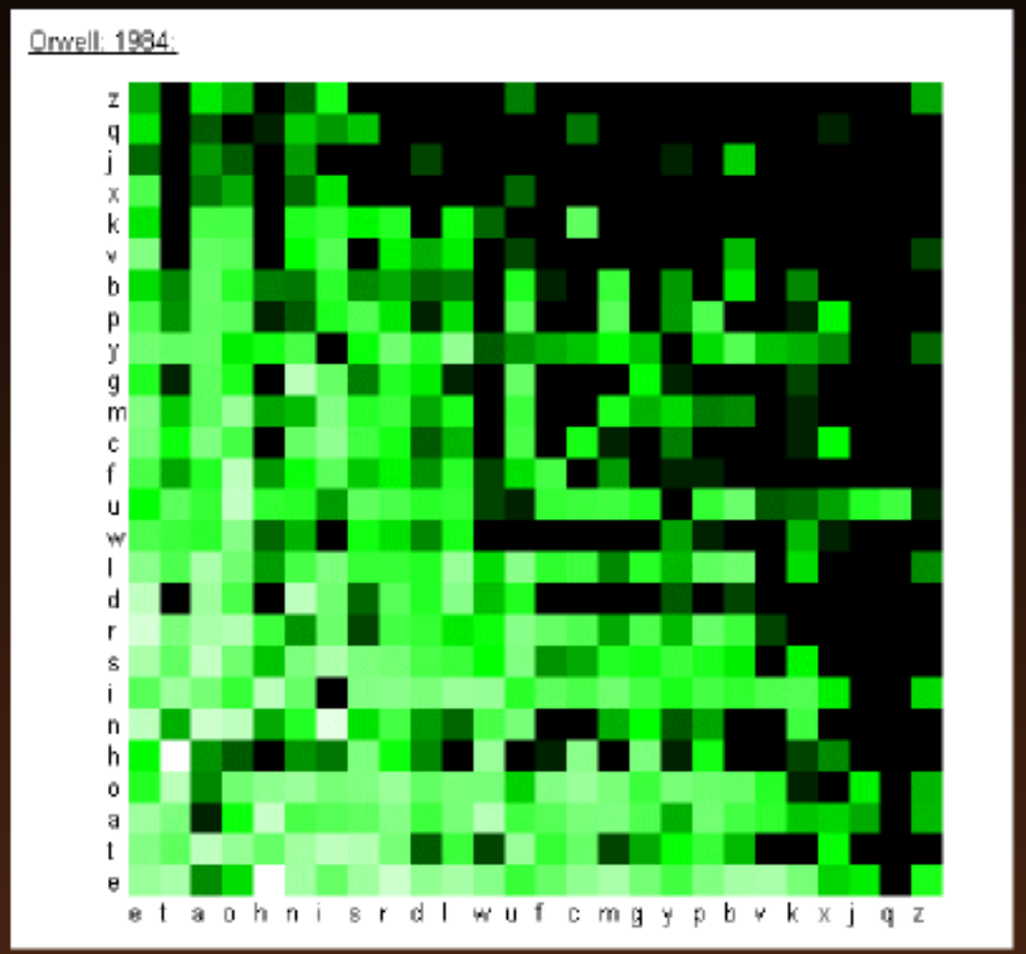
Simple statistical methods

Without the claim of the completeness:

- letters (however: glyph vs. grapheme):
  - verifying Zipf's Law (see earlier)
  - beginning and ending of words (thus, the whitespace is a special character)
- digraphs, trigraphs: see later on...
- groups of letters: Lempel – Ziv – Welch algorithm
- words:
  - length of the words
  - entropy: which words encode the more relevant information?
  - problem: similar words and/or words with the same meaning and having different form, Levenshtein-distance, ...
- larger units or blocks...

Probabilities of digraphs



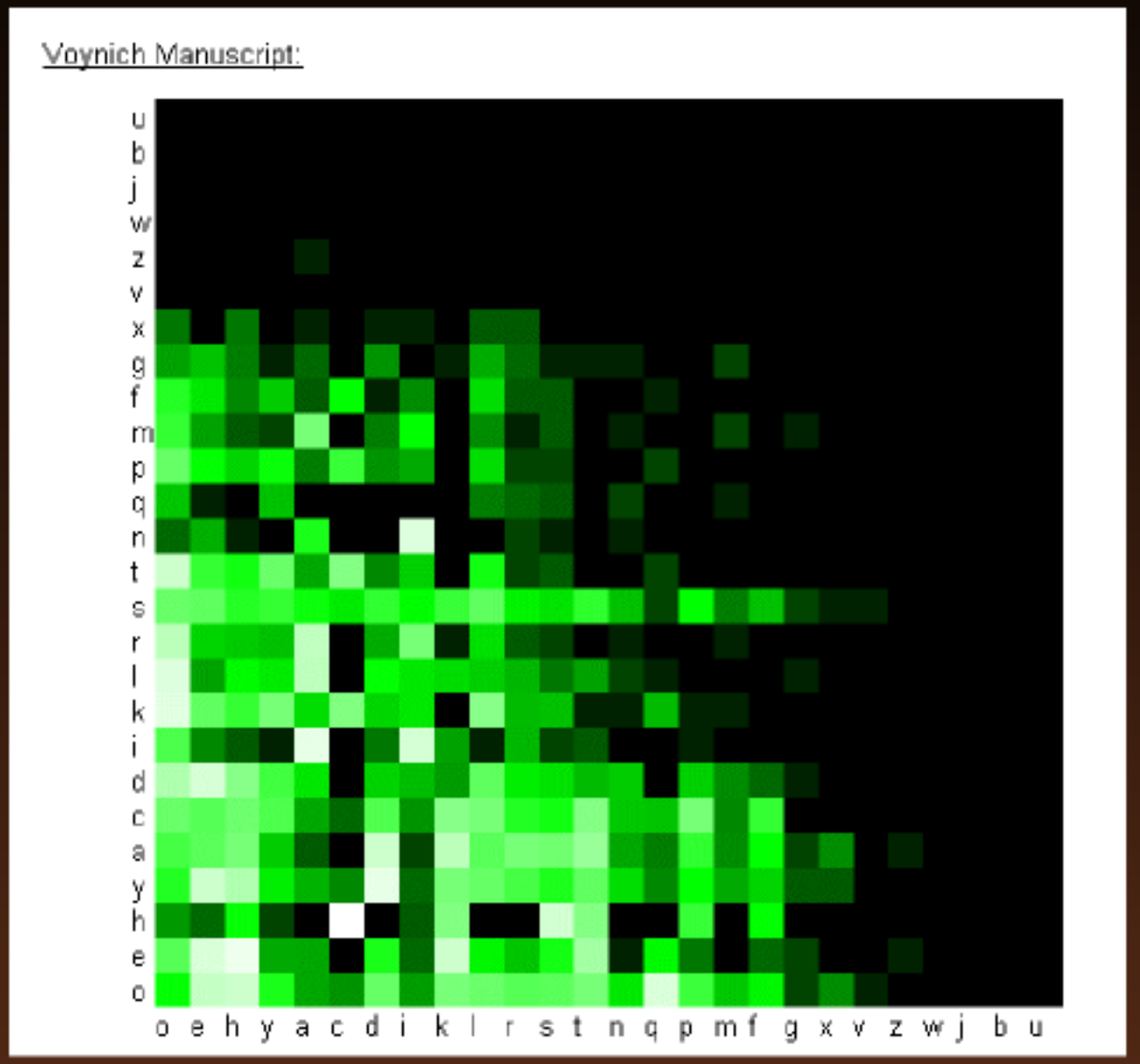


Conditional probabilities of digraphs

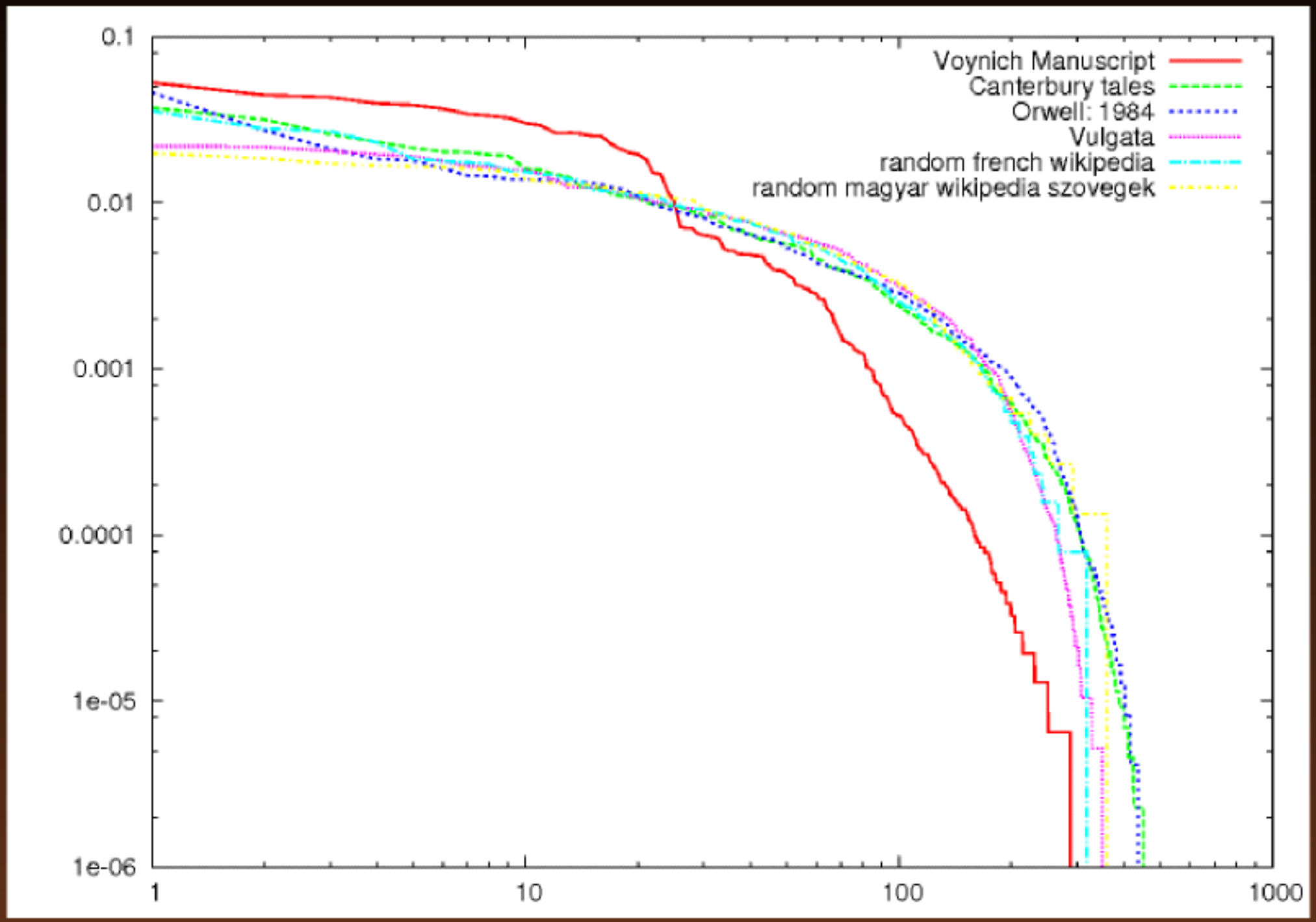
random magyar wikipedia szovegek:

	a	e	t	o	l	i	n	k	s	m	z	r	g	y	d	b	h	v	f	u	p	c	j	x	w	q
a	—	—	0.702	0.012	1.356	0.435	0.856	1.526	0.887	1.320	0.445	0.917	0.832	0.929	1.221	1.609	2.302	1.721	1.730	0.052	1.376	0.252	1.945	0.446	1.189	—
e	—	—	1.879	—	2.080	0.094	1.826	0.972	0.742	2.974	2.098	1.299	0.780	1.517	2.597	2.187	3.053	5.119	4.548	—	2.062	0.328	4.498	0.773	4.123	—
t	1.048	1.849	2.448	1.319	1.438	0.809	2.074	0.240	1.992	—	1.255	1.197	0.099	0.121	1.390	—	0.299	—	0.231	1.156	0.552	—	0.715	—	—	—
o	—	0.041	2.225	—	2.051	0.356	0.409	2.857	1.385	1.498	1.813	2.344	2.752	1.751	2.181	1.647	5.259	3.725	3.818	—	2.187	2.574	0.944	—	—	—
l	1.669	3.121	0.259	3.608	1.688	1.012	0.091	0.202	0.462	0.111	0.058	0.242	0.146	0.255	—	0.845	0.292	—	—	5.093	5.063	—	0.350	—	—	—
i	0.489	0.118	1.408	—	1.143	0.504	1.184	2.259	0.321	3.528	0.840	2.652	0.531	1.843	3.190	1.467	1.370	3.152	2.123	—	1.688	0.426	—	—	4.009	—
n	1.179	1.924	0.093	3.926	0.045	4.073	1.126	0.054	0.055	0.120	0.062	0.718	0.157	0.953	0.412	—	—	—	—	1.462	—	—	0.377	—	—	—
k	1.292	2.188	—	2.540	0.101	2.416	0.217	1.386	0.368	0.066	0.275	2.967	0.087	0.106	—	—	—	—	—	0.811	0.242	—	—	—	—	—
s	0.571	1.649	0.035	2.770	0.205	4.918	0.110	0.061	0.249	—	0.210	0.810	0.177	—	0.310	—	—	—	—	5.572	0.492	11.922	—	—	—	—
m	1.114	2.319	—	2.597	0.500	0.463	0.060	0.266	0.337	0.365	—	0.478	0.287	—	0.168	—	0.192	—	—	1.786	—	6.448	—	—	—	—
z	3.752	1.481	—	0.881	—	0.360	0.248	0.138	6.719	0.227	0.079	0.579	0.099	—	—	—	—	—	—	—	—	—	—	—	—	—
r	1.516	2.825	0.743	6.215	0.061	0.505	—	0.507	0.147	0.080	—	0.348	0.314	0.508	0.183	1.415	1.259	—	1.462	1.462	1.745	—	—	—	—	—
g	1.350	4.800	—	1.769	0.219	1.366	0.941	—	—	0.096	—	0.523	—	—	—	—	—	—	—	6.738	—	—	—	—	—	—
y	—	0.047	0.060	—	4.865	—	5.049	—	—	—	—	0.127	22.592	—	—	—	0.306	—	—	—	—	—	—	—	—	—
d	1.537	2.119	—	0.803	0.510	0.266	3.846	—	—	—	0.174	0.549	0.220	—	1.157	0.425	—	—	—	6.157	—	—	10.572	—	—	—
b	1.218	0.528	0.384	0.760	0.141	0.880	1.364	0.504	1.198	2.592	—	1.011	0.486	0.295	—	11.270	—	—	—	0.566	—	0.682	—	—	—	—
h	0.496	0.548	1.095	0.657	0.585	0.457	0.157	0.349	0.533	0.192	—	0.629	0.756	1.225	—	—	0.506	—	—	—	0.701	6.366	—	—	—	—
v	0.148	1.109	2.278	0.287	1.592	1.161	0.171	0.190	0.581	—	0.869	0.914	0.275	—	0.962	—	—	1.201	1.281	—	—	0.771	—	—	—	—
f	0.315	0.364	—	0.764	0.849	0.884	—	—	—	0.223	0.232	—	0.586	0.356	—	—	—	—	0.683	0.683	—	—	—	—	—	—
u	0.419	0.182	1.619	0.153	1.019	0.177	0.548	4.055	0.206	0.446	—	0.731	0.586	7.116	8.722	6.795	1.176	—	—	—	3.258	0.822	8.442	—	—	—
p	1.752	1.845	—	0.547	0.608	0.211	—	—	0.492	7.190	—	—	—	0.424	0.612	—	—	—	—	0.815	—	—	—	—	—	—
c	0.316	0.547	—	0.919	0.409	0.639	1.760	—	0.497	—	—	3.814	—	—	—	—	—	—	0.822	0.822	1.961	1.978	—	—	—	—
j	1.405	1.499	0.953	—	1.749	—	0.377	0.836	—	—	—	1.005	1.811	—	4.229	1.167	—	—	—	—	3.357	—	—	—	—	—
x	—	0.773	—	2.597	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
w	—	4.123	2.621	—	—	—	—	—	—	—	—	—	—	—	11.630	—	—	—	—	—	—	—	—	—	—	—
q	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—

Probability densities of digraphs – Voynich



Cumulative distribution of digraphs



The calendar

File: voynich-"dar.txt Line 1 Col 0 2780 bytes 100%

# PSC	# ARI	# TAU	# GEM	# CNC	# LEO	# VIR	# LIB	# SCO	# SGR
otaral	otalaiin	sholshdy	otal	yfary	shaleey	opals	okeey	okeos	otedy
otalar	oteoalols araly	ofaefom	salal	osaiinsal	oky	okeeom	oiny	okary	ched
otalam	oteoeey otal okealar	otalody	okam	ytoar shar	chcthey	olkar	oeey	okeeody	oty
dalaram	otear araydy	otalaiin	otalshy	octho	laly	oeedaly	okedy	oyteedy	?
otaram	okoly	otar shor	okaldy	aral	oeedy	cheos	ypa(?)	oky	okal
oteosal	opchey sal	chfaly	chosar	oletal	odainar	ofcheey	okalain	chefy	ykey
salols	otaka(??)	okolar	otam	opalal	okary	yteedy	oeeody	otal	ykchy
okaldal	okalal	otchody	ainaly	ytainal	ceosain	oraiin	okeal	shek	okeod
ykolaiin	otaly	alcphy	okarcham	oseesaiin	oteesed	okeosas	oeealy	ar	ykey
(??)am	oalcheg	otaiin	okal	?	?	okain	okydy	kar	opal
okyody	oteeodals	okaraiin	okaly	?	?	ockhol	oees	otaly	sheol
otyar	okolshy	okararaly	okal	?	?	oeedy	okeeoly	okeody	odees
okaly	otshshdy	oplaramdain	okeey ary	ykairaiin ainal	oreey	okeosam	odal	oteedyl	eshy
otody	otal ypsharal	opalarar	oteeary	okalar	ykam	ainom(?)	aiinod	ckhedy	ykeor
otald	otalchy tar amdy	char orom	otaindy	orary	okey	oshesy	odchdy	okeey	ofals
otaldar	otolchdy	otchoshy	okaincham	olaiin olackhy	otcheody	okeoly	oteod	okeedy	ykeody
okody	otoloaram	otchdal	okealy	olfsheoral	okyeshey	cheoothy	yteod	ykeeory	oteody
opysam	oteeol	okeeyary	otaror	aralkam	(?)ofaiin	?	okchody	oteeosy	ypaiin
chckhey	otolchd	otainy	otaly	olalsy	okaly	sheyoky	okal	shekal	yfaiiny
otaly	otaldar	ofaralar	orary	oraiinam	otaly	oteodas	ytaly	oteedyg	ofcheesy
otalarar	oteoaldy	oaiinarary	-	osarshee(?)	ofalals	ofchdysd	oechos	otedal	oraiiny
otaldy	oteolar	okalam	otyd	ota(?)aiin	? ocfhy	oteood	oteofy	chdy	ykeeody
okeoly	okeoaly	ytalshdy	otalam	opaiinor alain	daiinchepy	yteod	cheoepy	dalshey	okeos
okydy	otaleky	chor alep	otaraldy	ypain aloly	oteeyday	okeody	ykeedy	opaiin	ykeear
otees	opalshar	otaraldy	okalar	otchydaiin	?	okeoldy	ochoty	okeos	qokeoly
otalalm	cheary	otsiin otain	opchdady	oeeodaiin	?	ykeeas	octhy	otey	okeody
okady	oteotey sary	otalep as aiinam	oklaindy	ofsholdy	?	opaiin(?)	oteoly	otaly	okal
otar	otalaly	ocholsharam	okaram	?	?	choldy	okeoly	chockhy	oteody
oty	oteos arar	oshodady	okainy	ykolairol	?	ochdey	oeees	okedy	oteody
	okldam	chdaiindaiiny	okealar	olkalain	?	oeeodaiin	ocfhshy	yteeody	cheody

1Help 2Wrap 3Quit 4Hex 5Line 6RxSrch 7Search 8Raw 9UnForm 10Quit

f70r, . . . , f73v

Large scale structure

- It can safely be assumed: 4 types of handwriting, seems to be continuous (so „slow" encoding can be ruled out: it was known what has to be written)
- The cumulative distribution of the digraph clearly shows that there were more than one author or transcripter. Similar studies can be conducted on modern text as well (see e.g. Orwell: 1984, the book of Goldstein).
- A typical example: the ending 'edy' can only be found after f31r. If we check the quires, such relations are even more obvious.

Conspiracy theories

- It is a pure hoax (see also Rudolph II.);
- A document for a parallel universe;
- Diary of a time traveller;
- Documentation of alien visits;
- Fortune telling;
- It is not by chance, that 37 pages are missing...

(http://konteo.blogrepublik.eu/2011/02/15/a-voynich-kezirat/, in Hungarian)

Thank you