



¶ *Psalmus ex mōdō: uñ h̄zyōſ. A h̄mōdēcōdē ſc̄p̄t̄r̄c̄ȳm
x̄urx̄gzm̄s: poſūlāc̄: uñ x̄ijh̄ēp̄ua: m̄zv̄ | Δwph̄c̄p̄q̄ | n̄x̄ūt̄
h̄ui+ = n̄lāc̄r̄ = l̄v̄ngch̄c̄: oꝝ p̄iðr̄m̄: l̄on Δuñ x̄j̄t̄+ūc̄p̄āl̄f̄
m̄iñzue | h̄oçip̄a3n̄s̄x̄d̄ | p̄s̄z̄d̄+l̄r̄uñp̄c̄āȳr̄eññh̄j̄+t̄l̄
aðh̄m̄q̄zurx̄iñr̄oſuñh̄p̄r̄c̄: iñi = ſw̄om̄ȳd̄n̄n̄c̄: uñ Δr̄x̄ḡ
h̄m̄ l̄n̄d̄t̄iñr̄d̄l̄uñp̄ořeſač̄: t̄n̄x̄p̄m̄z̄ȳ | l̄ðāzr̄ořiñr̄u
p̄h̄c̄: + h̄iñc̄d̄ = ḡr̄aðr̄m̄: alp̄c̄z̄ηp̄uñz̄iñr̄iñr̄m̄h̄ož̄ež̄+c̄uñr̄
h̄s̄iñr̄uñk̄ořiñp̄t̄m̄t̄h̄uñh̄iñr̄f̄.*

N^o 3

Wieder entzündet sich gallertende Eiter und
an Händen, Füßen, Schultern, allen Körpers
entzündet sich die Eiter. Und wenn es
heute Abend nicht geheilte werden sollten
so wird es morgen nicht geheilte werden.
Sobald es morgen nicht geheilte werden
so wird es morgen nicht geheilte werden.



Franchi allora
scrisse questo
canto per la sua
morte
e così si scrisse
che il poeta
di cui si parla

¶ Verba eis deo dñe et eis deo dñe galland dñe eis
et eis eis galland dñe galland eis deo dñe eis eis
galland dñe galland dñe galland eis deo dñe eis
eis eis galland eis deo dñe galland eis galland dñe
dñe eis galland eis deo dñe galland eis galland dñe

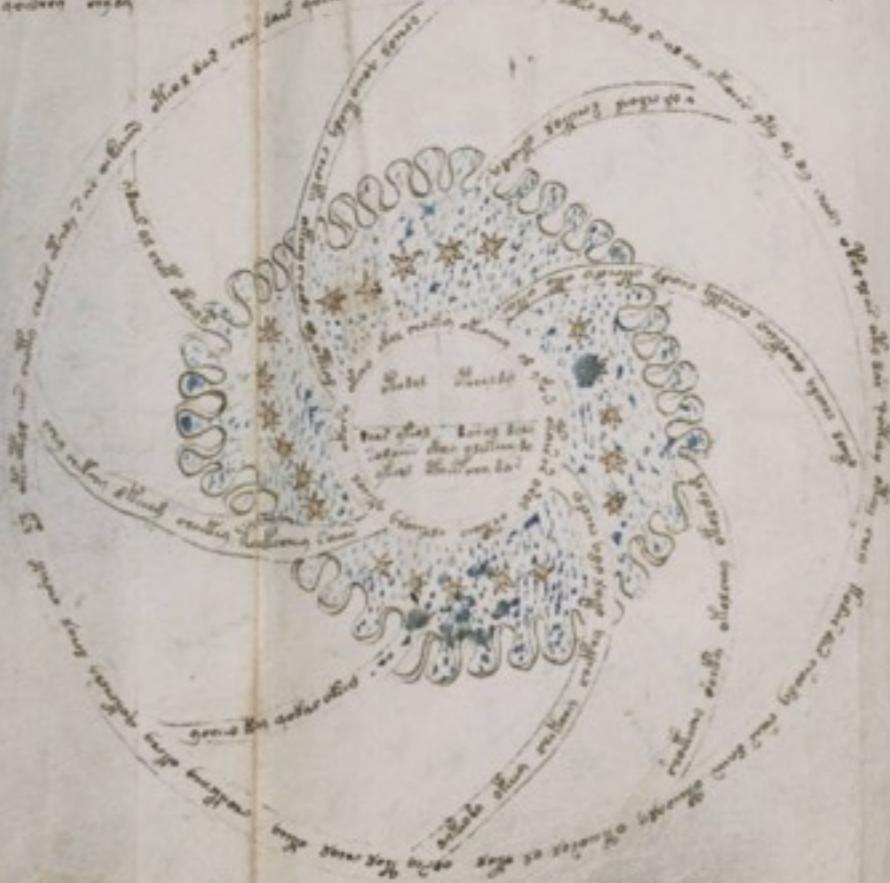


¶ Þótt er ekki fáskar að búa til með þessum dæmi sínar
þann frábæra vallans vestur millan óhvar ófarið hafið
Hobnum meðgötuðu gosar eftir af óllum ófarið ófari
þann frábæra vallan ófarið.

Þessi gosar eru mikil með eftir ófarið og eru fáir
a gosar af ófarið ófarið eftir ófarið ófarið ófarið ófarið
þann frábæra vallans vestur millan óhvar ófarið hafið
Hobnum meðgötuðu gosar eftir ófarið og eru ófarið
þann frábæra vallan ófarið.



Truly vifing amys des houres vnties foyr england entredis this quodding
vngloues vng vng vng blaynes vng dene vng yelde and vng vng vng
the vngloue vng
vng vng vng vng vng vng vng vng vng vng vng vng vng vng vng vng vng



63
Truly vifing vng
vng vng vng vng vng vng vng vng vng vng vng vng vng vng vng vng vng
vng vng vng vng vng vng vng vng vng vng vng vng vng vng vng vng vng
vng vng vng vng vng vng vng vng vng vng vng vng vng vng vng vng vng



Decipherment

Decipherment

Most material from Kevin Knight (USC/ISI)

Decipherment

Most material from Kevin Knight (USC/ISI)

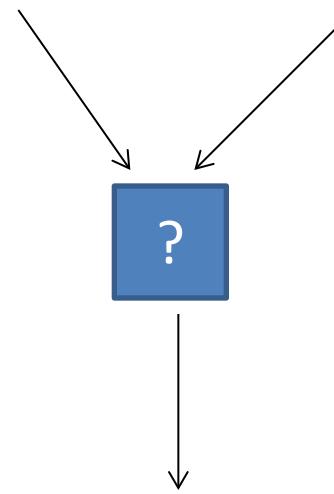
Bad timing and terrible jokes are all mine, though.

Learn Translation Knowledge from Non-Parallel Text?

Parallel text



English text

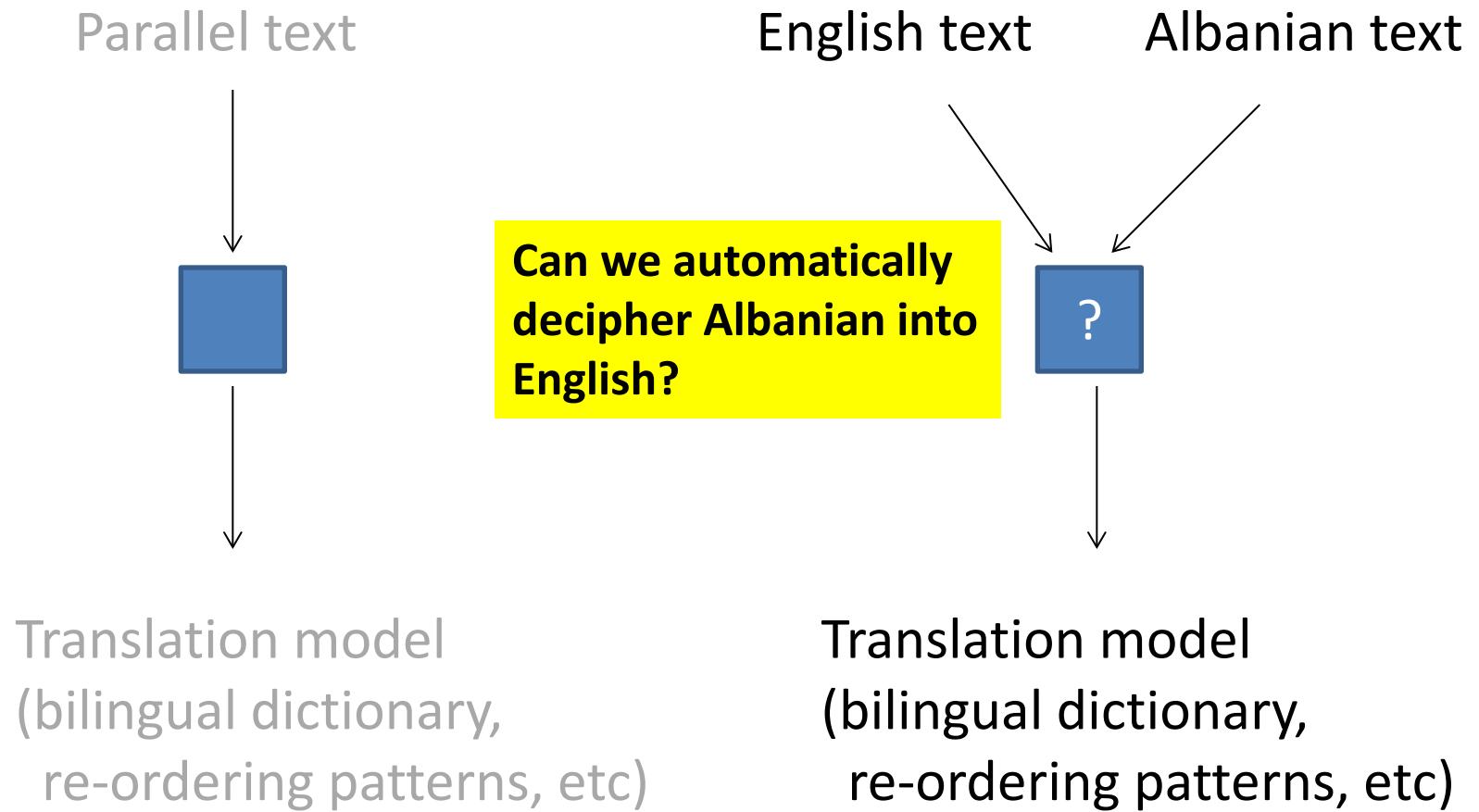


Albanian text

Translation model
(bilingual dictionary,
re-ordering patterns, etc)

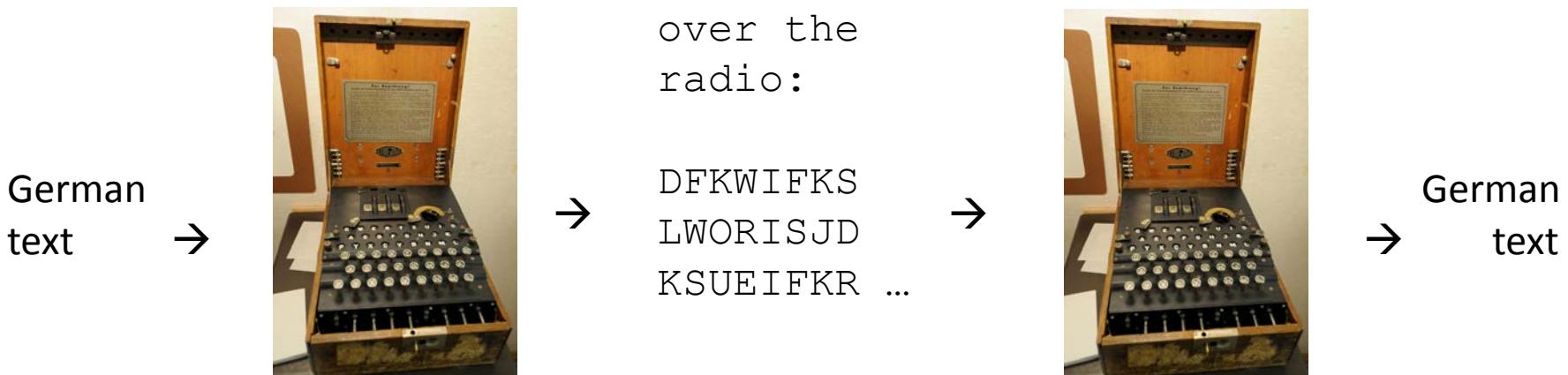
Translation model
(bilingual dictionary,
re-ordering patterns, etc)

Learn Translation Knowledge from Non-Parallel Text?



“First NLP Task Ever” (1930s-40s)

Breaking the German Enigma Cipher



input (intercepted ciphertext) :
output (plaintext) :

DFKWIFKSLWORISJDKSUEIFKR ...
VASISTDASHERRCAPITANRICH ...

You can think of this like a multi-class tagging problem.
Each letter of ciphertext can get one of 26 tags (plaintext letters A...Z).

“First NLP Task Ever” (1930s-40s) Breaking the German Enigma Cipher

Substitution system

$N \rightarrow J$

Substitution table **changes** with every keystroke:

$NNN \rightarrow JTE$

Flattens out ciphertext letter distributions.

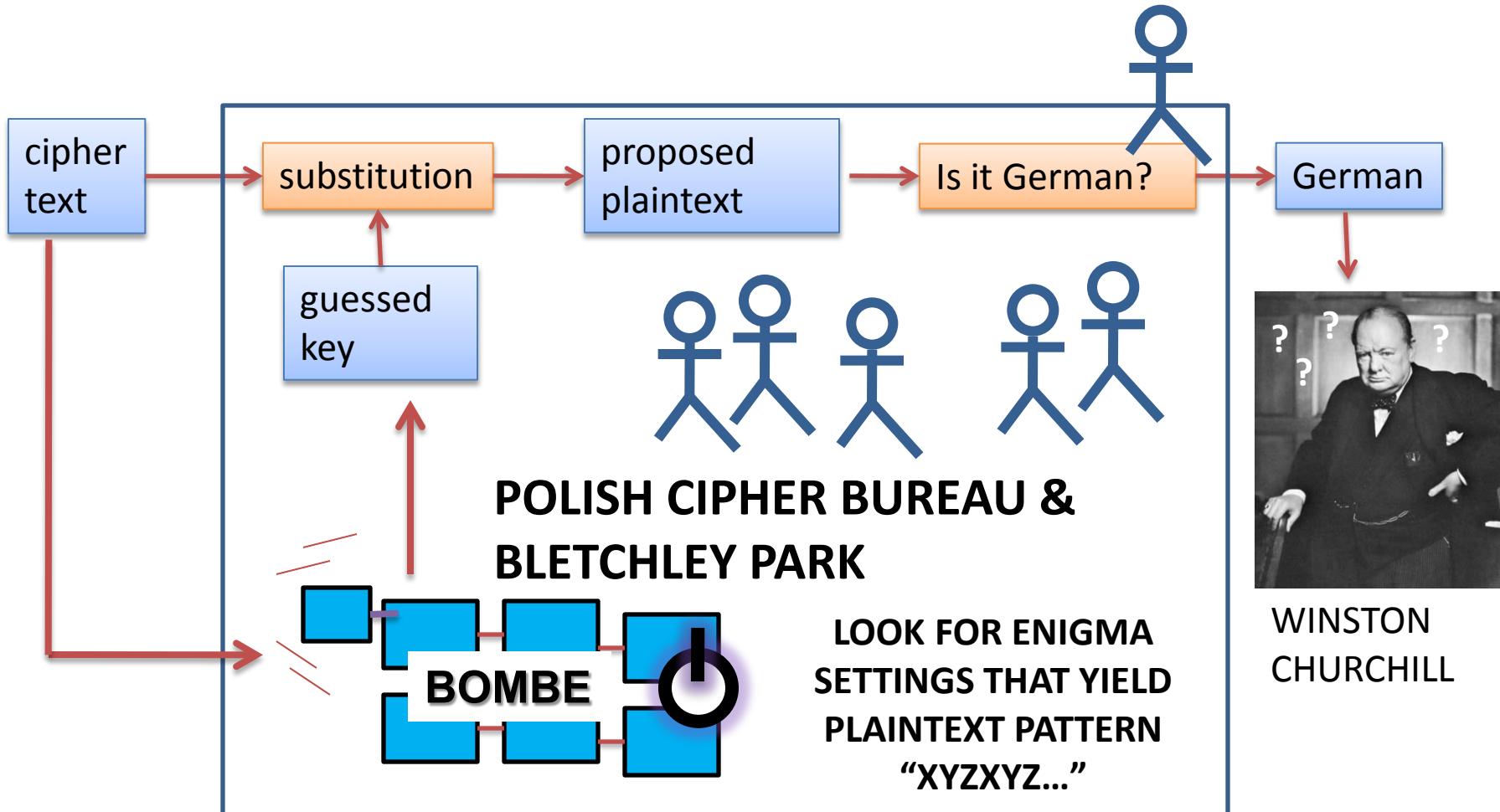


Secret key =
initial rotor
ordering and
settings

Reversible behavior

$NNN \rightarrow JTE \rightarrow NNN$

Breaking Enigma



Letter Substitution Cipher

- Encipherment key:

PLAIN: ABCDEFGHIJKLMNOPQRSTUVWXYZ

CIPHER: PLOKMIJNUHBYGVTFCRDXESZAQW

- Plaintext: **HELLO WORLD . . .**
- Ciphertext: **NMYYT ZTRYK . . .**
- Key itself doesn't change: "simple substitution"
- What key, if applied to the ciphertext, would yield sensible plaintext?

KDCY LQZKTLJKX CY MDBCYJQL: "TR

HYD FKXC, FQ MKX RLQQIQ HYDL

MKL DXCTW RDCDLQ JQMNKXTMB

PTBMYEQL K FKH CY LQZKTL TC."

KDCY LQZKTLJKX CY MDBCYJQL: "TR

HYD FKXC, FQ MKX RLQQIQ HYDL

MKL DXCTW RDCDLQ JQMNKXTMB

PTBMYEQL K FKH CY LQZKTL TC."

A	
B	3
C	8
D	7
E	1
F	3
G	
H	3
I	1
J	3
K	10
L	10
M	6
N	1
O	
P	1
Q	10
R	3
S	
T	7
U	
V	
W	1
X	5
Y	7
Z	2

KDCY LQZKTLJKX CY MDBCYJQL: "TR

HYD FKXC, FQ MKX RLQQIQ HYDL

MKL DXCTW RDCDLQ JQMNKXTMB

PTBMYEQL K FKH CY LQZKTL TC."

A	
B	3
C	8
D	7
E	1
F	3
G	
H	3
I	1
J	3
K	10
L	10
M	6
N	1
O	
P	1
Q	10
R	3
S	
T	7
U	
V	
W	1
X	5
Y	7
Z	2

KDCY LQZKTLJKX CY MDBCYJQL: "TR

HYD FKXC, FQ MKX RLQQIQ HYDL

MKL DXCTW RDCDLQ JQMNKXTMB

PTBMYEQL K FKH CY LQZKTL TC."

A	
B	3
C	8
D	7
E	1
F	3
G	
H	3
I	1
J	3
K	10
L	10
M	6
N	1
O	
P	1
Q	10
R	3
S	
T	7
U	
V	
W	1
X	5
Y	7
Z	2

#

.

.

V

##

#

.

.

.

V

#####

.

.

V

###

.

.

V

###

V

.

a . a . a

KDCY LQZKTLJKX CY MDBCYJQL: "TR

. . a . a . . .

HYD FKXC, FQ MKX RLQQIQ HYDL

a a

MKL DXCTW RDCDLQ JQMNKXTMB

. . a . a . a

PTBMYEQL K FKH CY LQZKTL TC."

A	
B	3
C	8
D	7
E	1
F	3
G	
H	3
I	1
J	3
K	10
L	10
M	6
N	1
O	
P	1
Q	10
R	3
S	
T	7
U	
V	
W	1
X	5
Y	7
Z	2

. # # # # V
V

. #
V
V

a e.a .a .e .

KDCY LQZKTLJKX CY MDBCYJQL: "TR

. .a .e a . ee.e .

HYD FKXC, FQ MKX RLQQIQ HYDL

a . . e .e .a

MKL DXCTW RDCDLQ JQMNKXTMB

. .e a .a. e.a

PTBMYEQL K FKH CY LQZKTL TC."

didn't create "ae"

A	
B	3
C	8
D	7 #
E	1 .
F	3 .
G	
H	3 .
I	1 .
J	3 .
K	10 ##### V
L	10 ##
M	6 #
N	1 .
O	
P	1 .
Q	10 ##### V
R	3 .
S	
T	### V
U	
V	
W	1 .
X	5
Y	7 ### V
Z	2 .

a e .ao .a .e o .

KDCY LQZKTLJKX CY MDBCYJQL: "TR

. .a .e a . ee.e .

HYD FKXC, FQ MKX RLQQIQ HYDL

a o . . e .e .a o

MKL DXCTW RDCDLQ JQMNKXTMB

.o .e a .a. e .ao o

PTBMYEQL K FKH CY LQZKTL TC."

don't like "ao" – back up!

A	
B	3
C	8
D	7 #
E	1 .
F	3 .
G	
H	3 .
I	1 .
J	3 .
K	10 ##### V
L	10 ##
M	6 #
N	1 .
O	
P	1 .
Q	10 ##### V
R	3 .
S	
T	### V
U	
V	
W	1 .
X	5
Y	7 ### V
Z	2 .

auto repairman to customer: if

KDCY LQZKTLJKX CY MDBCYJQL: "TR

you wait we can freeze your

HYD FKXC, FQ MKX RLQQIQ HYDL

car until future mechanics

MKL DXCTW RDCDLQ JQMNKXTMB

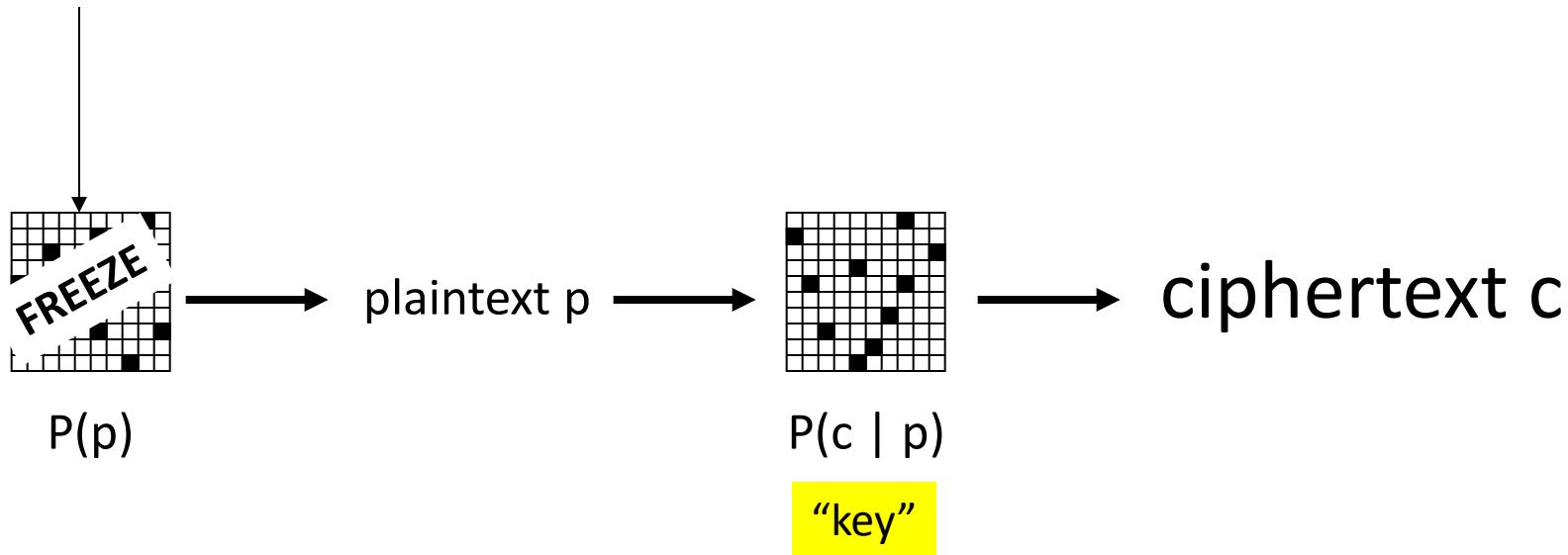
discover a way to repair it

PTBMYEQL K FKH CY LQZKTL TC."

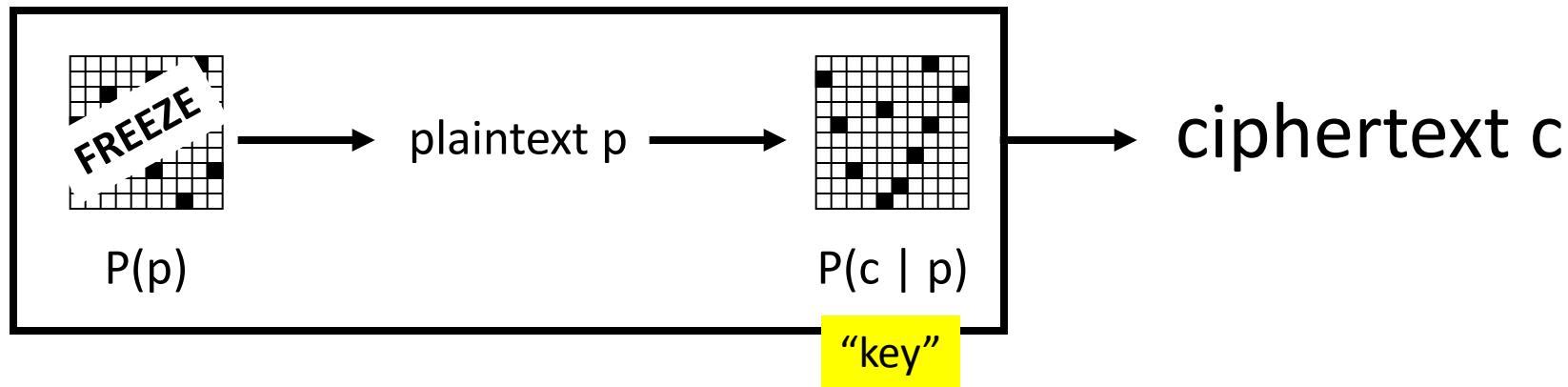
A		
B	3	
C	8	
D	7	#
E	1	.
F	3	.
G		
H	3	.
I	1	.
J	3	.
K	10	##### V
L	10	##
M	6	#
N	1	.
O		
P	1	.
Q	10	##### V
R	3	.
S		
T	7	### V
U		
V		
W	1	.
X	5	
Y	6	### V
Z	2	.

Letter Substitution Cipher

plaintext samples,
unrelated to ciphertext



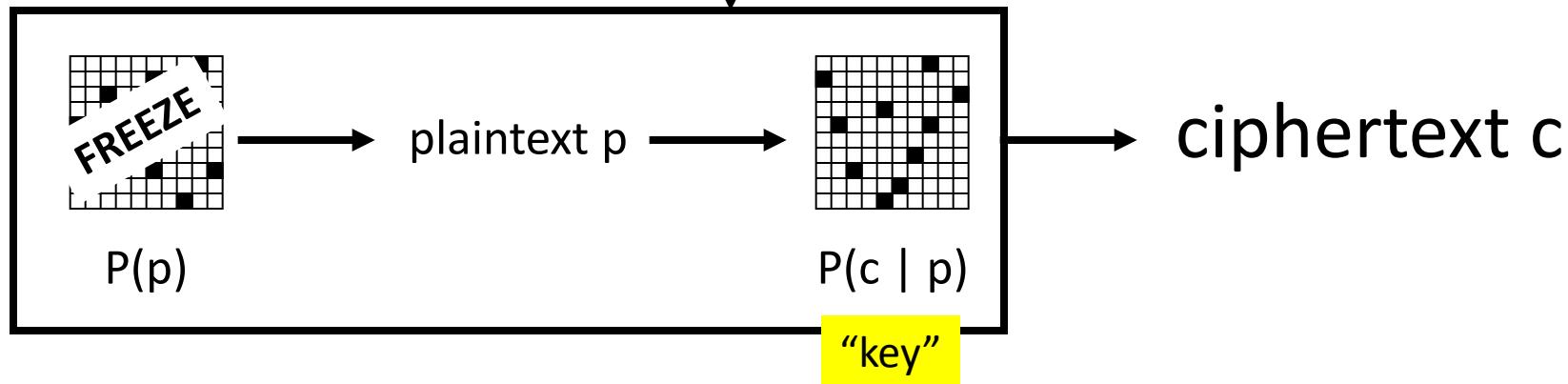
Letter Substitution Cipher



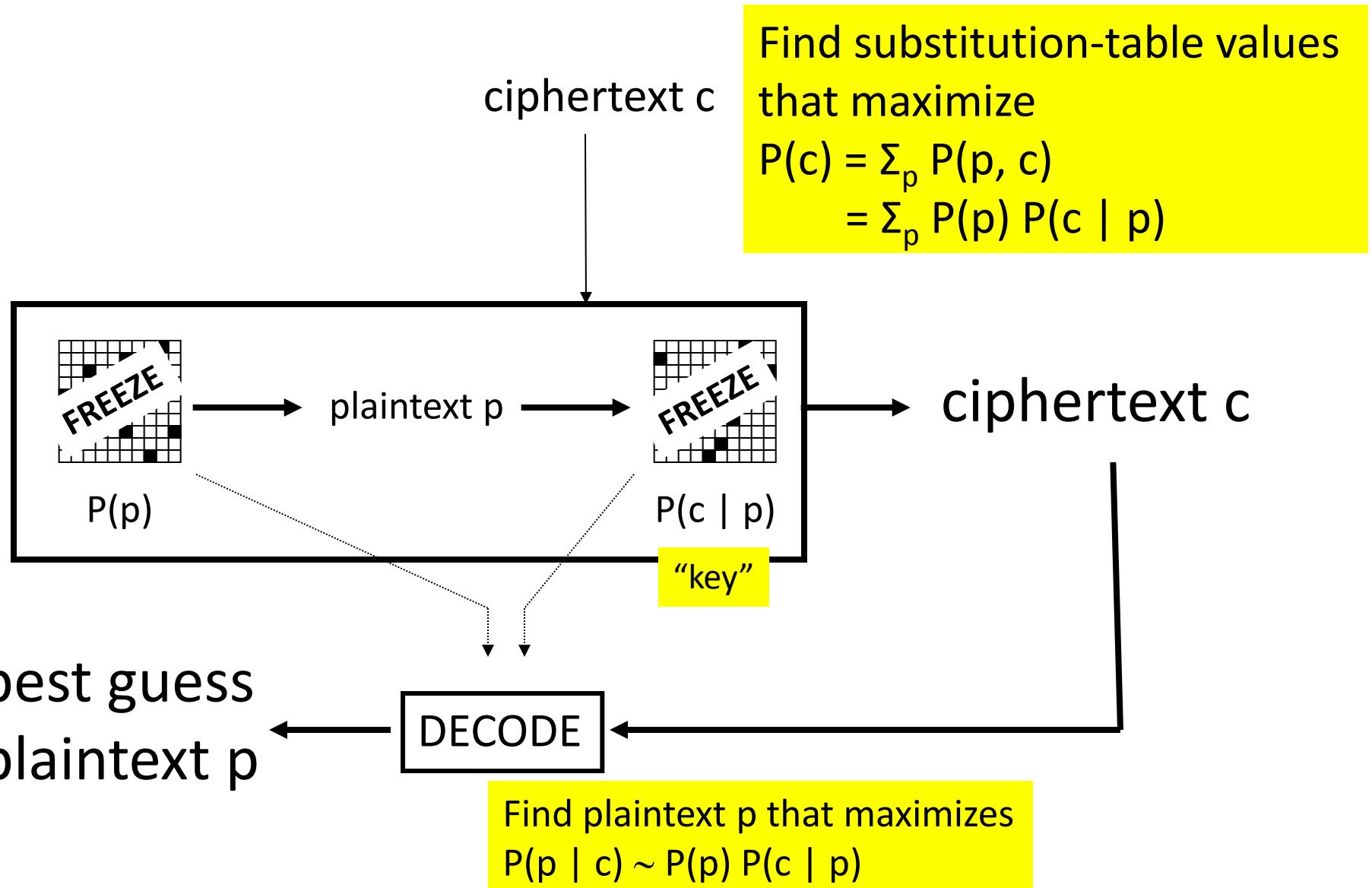
Letter Substitution Cipher

ciphertext c

Find substitution-table values
that maximize
 $P(c) = \sum_p P(p, c)$
 $= \sum_p P(p) P(c | p)$



Letter Substitution Cipher

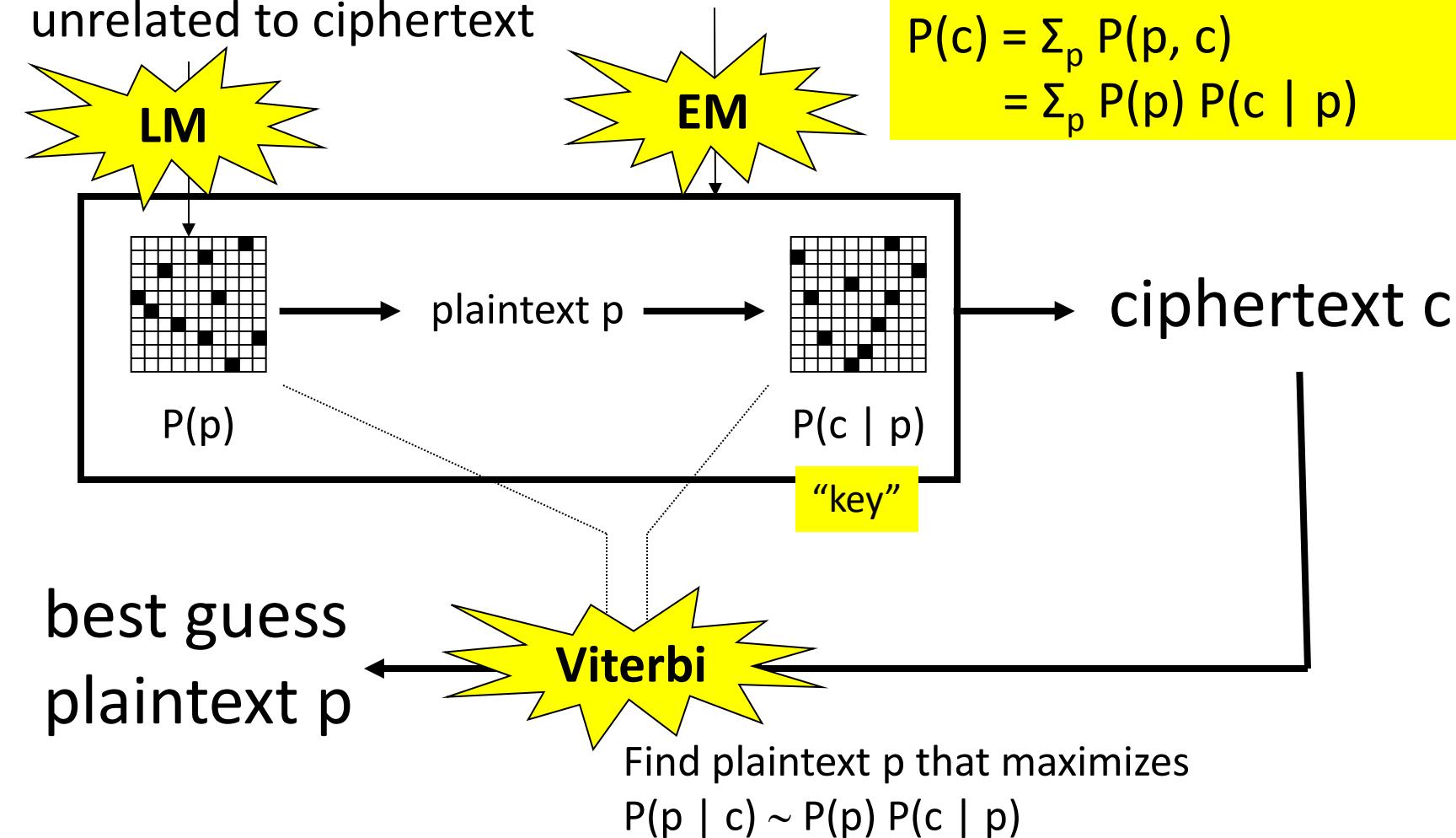


Letter Substitution Cipher

plaintext samples,
unrelated to ciphertext

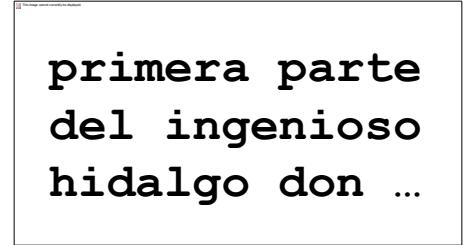
ciphertext c

Find substitution-table values
that maximize
 $P(c) = \sum_p P(p, c)$
 $= \sum_p P(p) P(c | p)$



Phonetic Decipherment

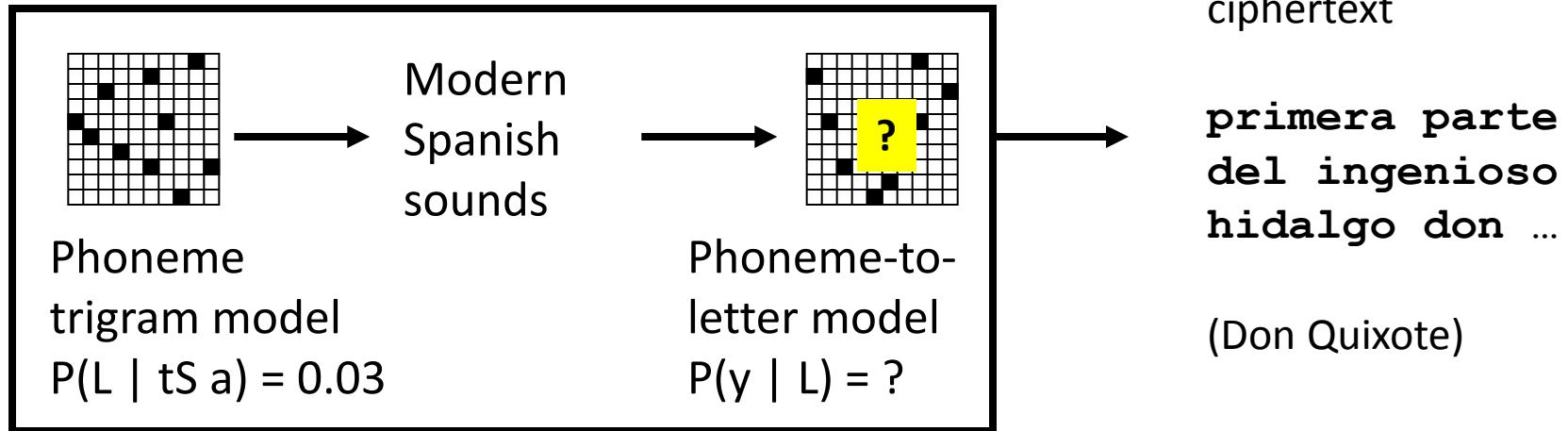
ciphertext



A screenshot of a text editor window titled "Untitled - Microsoft Word". The window contains the following text in Spanish:

primera parte
del ingenioso
hidalgo don ...

Phonetic Decipherment



26 sounds:

B, D, G, J (canyon),
L (yarn), T (thin), a,
b, d, e, f, g, i, k, l,
m, n, o, p , r,
rr (trilled), s,
t, tS, u, x (hat)



32 letters:

ñ, á, é, í, ó, ú,
a, b, c, d, e, f, g,
h, i, j, k, l, m, n,
o, p, q, r, s, t, u
v, w, x, y, z

EM approach = 93% accurate phonetic decipherment

What if Spoken Language Behind Script is Unknown?

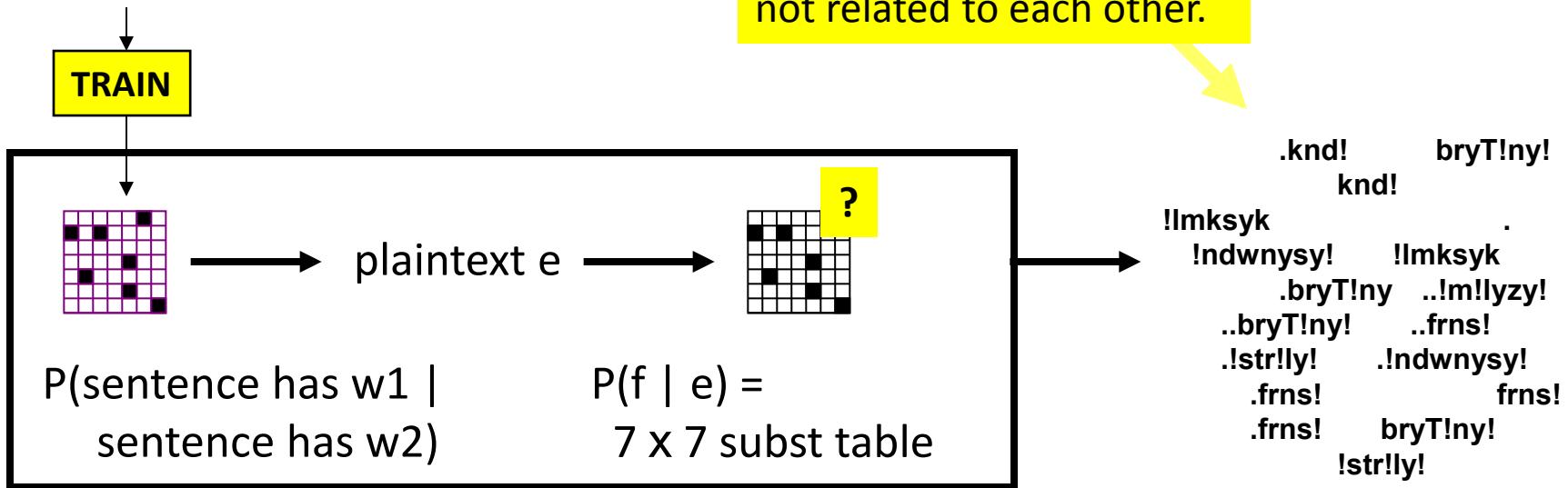
- Build a universal model $P(p)$ of human phoneme sequence production
 - human might generally say: K AH N AH R IY
 - human won't generally say: R T R K L K
- Find a $P(c | p)$ table
 - such that there is a decoding with a good universal $P(p)$ score
- Phoneme & syllable inventory
 - if z, then s
 - all have CV syllables; if VCC, then also VC
- Syllable sonority structure
 - dram, lomp, ? rdam, ? lopm
- Physiological preference constraints
 - tomp, tont, ? tomk, ? tonp

[Knight et al 06]

Word Substitution Cipher

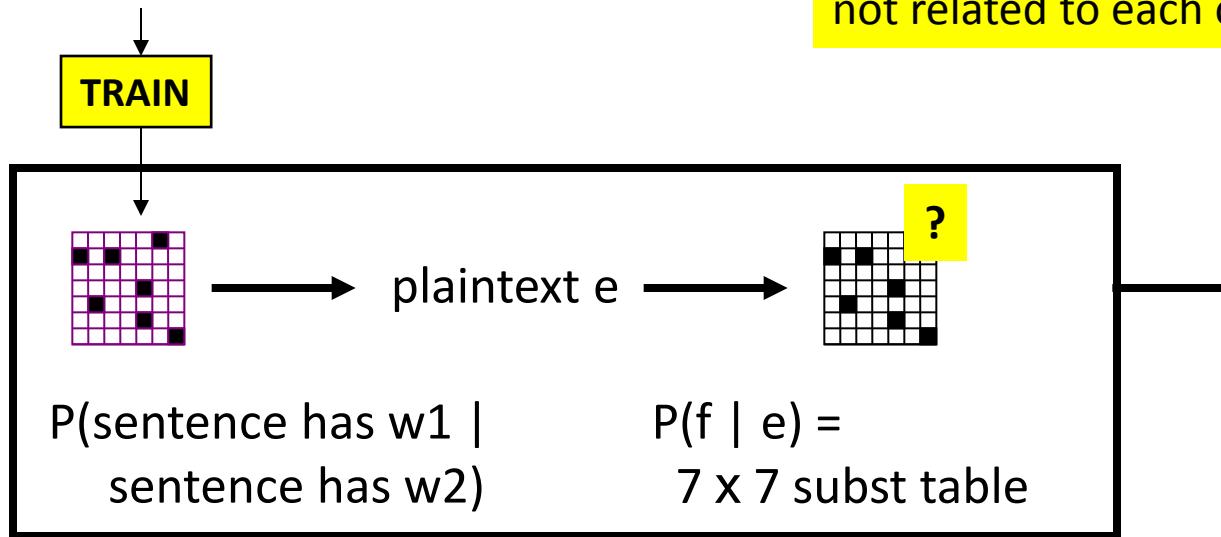
.France	Britain	Canada
Mexico	Indonesia	Malaysia
.Britain	..Canada	..Australia
..Britain	.France	.Indonesia
.Mexico	Australia	.France
Britain	.	.

Key Point: These texts are not related to each other.



Word Substitution Cipher

.France	Britain	Canada
Mexico	Indonesia	Malaysia
.Britain	..Canada	..Australia
..Britain	.France	.Indonesia
.Mexico	Australia	.France
Britain	.	.



.knd!	bryT!ny!
knd!	.
!lmksyk	!ndwnysy!
!lmksyk	!lmksyk
.bryT!ny	..!m!lyzy!
..bryT!ny!	..frns!
.!str!ly!	.!ndwnysy!
.frns!	frns!
.frns!	bryT!ny!
!str!ly!	.

Australia	\rightarrow	!str!ly!	(0.93)	!ndwnysy!	(0.03)	m!lyzy!	(0.02)
Britain	\rightarrow	bryT!ny!	(0.98)	!ndwnysy!	(0.01)	!str!ly!	(0.01)
Canada	\rightarrow	knd!	(0.57)	frns!	(0.33)	m!lyzy!	(0.06)
France	\rightarrow	frns!	(1.00)				
Indonesia	\rightarrow	!ndwnysy!	(1.00)				
Malaysia	\rightarrow	m!lyzy!	(0.93)	lmksyk	(0.07)		
Mexico	\rightarrow	!lmksyk	(0.91)	m!lyzy!	(0.07)		

[Knight et al 06]

!!@!m
!lywm
!lth!ny&
!!@!m !lm!Dy
Sfr
@!m
th!ny&
@!m 1992
@!m 1993
ywm
!!!sbw@ !lm!Dy
fy !ldqyq&
!lsn& !lj!ry&
!lsn&
!lsh=hr !lm!Dy
!lsh=hr !lj!ry
snw!t
sn&
=hdh! !!@!m
s!@&
!!@Sr
@!m 1991

Time Expressions

@!m 1990
w!lth!ny&
fy !lywm
mn !lsh=hr !lj!ry
!lqrn
!y!m
@!m!aN
!!s!@&
17 shb!T 1994
th!lth snw!t
dqyq&
=hdh=h !lsn&
ywmyn
mn !!@!m !lm!Dy
!lsn& !lmqbl&
fy !lsn&
kl ywm
fy !!@!m !lm!Dy

!!@Swr
=hdh! !lsh=hr
fy ywm
nys!n
!sbw@
=hdh=h !!!'y!m
qbl !y!m
fy !!@Sr
mn !lsn&
!lsnw!t
b@d ywm
!!y!m
13 nys!n 1994
!lth!ny& @sh!&
th!lth& ly!m
qbl !sbw@yn
fy !lywm !lt!ly
sh@b!n
tmwz
3 dhw !!Hj& 1414
fy shb!T !lm!Dy
qbl ywmyn

Time Expressions

< n > < n > * ??? 19 < n > < n >

9 Hzyr!n 1942	27 tmwz 1993	21 Hzyr!n 1967
8 tshrym !!!wl 1990	26 tmwz 1953	20 !'y!r 1990
7 k!nwn !!!wl 1993	26 shb!T 1993	20 tshrym !'wl 1983
6 !'y!r 1993	26 k!nwn !!!wl 1994	20 tshrym !!!wl 1921
6 !~Adh!r 1991	25 !ylwl 1926	1 !y!r 1994
5 shb!T 1950	24 !~Adh!r 1993	17 Hzyr!n 1972
4 Hzyr!n 1989	22 !ylwl 1957	16 !ylwl 1919
30 !~Adh!r 1944	22 tshrym !!!wl 1948	16 Hzyr!n 1984
29 !y!r 1945	22 tmwz 1952	16 !~Ab 1929
29 !~Adh!r 1993	21 !y!r 1994	
28 k!nwn !!!wl 1994	21 k!nwn !!!wl 1988	

Time Expressions

< n > Hzyr!n < n >

13	4 Hzyr!n 1967	2	fy 30 Hzyr!n 1995
12	fy 12 Hzyr!n 1993	2	fy 18 Hzyr!n 1994
7	5 Hzyr!n 1967	2	fy 14 Hzyr!n 1993
6	fy 30 Hzyr!n 1989	2	fy 14 Hzyr!n 1991
6	30 Hzyr!n 1989	2	fy 12 Hzyr!n 1990
4	fy 30 Hzyr!n 1994	2	7 Hzyr!n 1994
4	fy 30 Hzyr!n 1993	2	6 Hzyr!n 1941
3	fy 19 Hzyr!n 1967	2	26 Hzyr!n 1994
2	ywm 30 Hzyr!n 1989	2	21 Hzyr!n 1994
2	w 6 Hzyr!n 1994	2	1 Hzyr!n 1994
2	qbl 5 Hzyr!n 1967	2	19 Hzyr!n 1965
2	fy 9 Hzyr!n 1967	2	18 Hzyr!n 1994
2	fy 7 Hzyr!n 1981	2	18 Hzyr!n 1940
2	fy 6 Hzyr!n 1994	2	12 Hzyr!n 1993
2	fy 5 Hzyr!n 1967	2	11 Hzyr!n 1994

Time Expressions

<n> Hzyr!n <n>

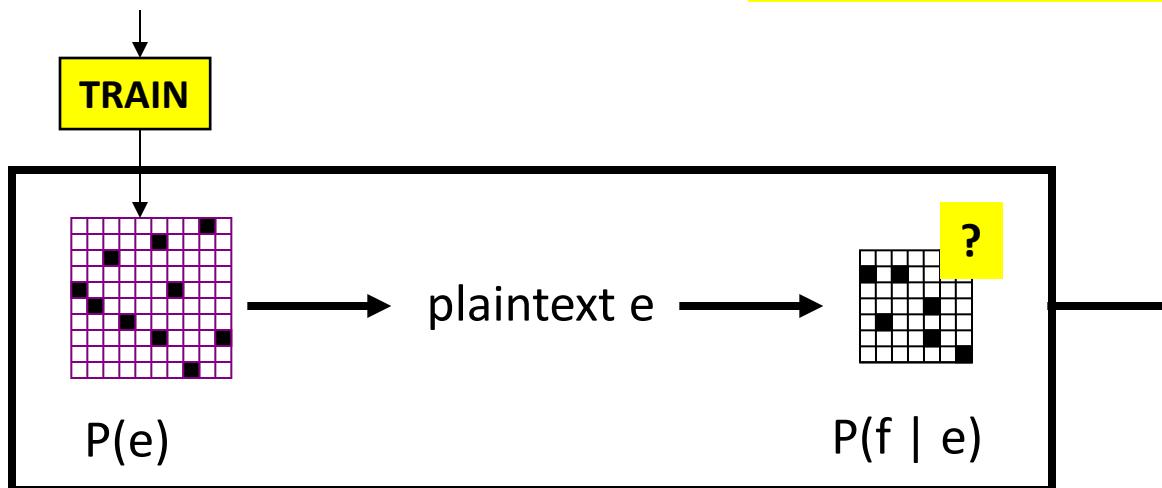
13	4 Hzyr!n 1967
12	fy 12 Hzyr!n 1993
7	5 Hzyr!n 1967
6	fy 30 Hzyr!n 1989
6	30 Hzyr!n 1989
4	fy 30 Hzyr!n 1994
4	fy 30 Hzyr!n 1993
3	fy 19 Hzyr!n 1967
2	ywm 30 Hzyr!n 1989
2	w 6 Hzyr!n 1994
2	qbl 5 Hzyr!n 1967
2	fy 9 Hzyr!n 1967
2	fy 7 Hzyr!n 1981
2	fy 6 Hzyr!n 1994
2	fy 5 Hzyr!n 1967

Search query	Documents
January 4, 1967	8040
February 4, 1967	9270
March 4, 1967	10700
April 4, 1967	21800
May 4, 1967	14000
June 4, 1967	39300
July 4, 1967	12600
August 4, 1967	7970
September 4, 1967	7390
October 4, 1967	8800
November 4, 1967	6560
December 4, 1967	9770

Foreign Language as a Cipher

BAGHDAD, Iraq (CNN) -- Six bombings killed at least 54 Iraqis and wounded 96 others Wednesday, including 20 civilians who died as they lined up to join the Iraqi army in Hawija when a suicide bomber detonated explosives hidden under his clothing, Iraqi officials said. That attack in the town about 130 miles (209 kilometers) north of Baghdad also wounded 30 Iraqis, said Iraqi army Lt. Col. Khalil al-Zawbai. A car bombing in Saddam Hussein's ancestral homeland of Tikrit also killed 30 Iraqis and wounded another 40, Iraqi officials said. The Tikrit explosion

Key Point: These texts are not related to each other.



رفض رئيس السلطة الفلسطينية محمود عباس مجددا تصريحات وزير الخارجية الإسرائيلي سيلفان شالوم التي قال فيها إنه يتمنى على إسرائيل إعادة النظر في انسحابها من غزة، المقرر أن يتم الصيف المقبل إذا فازت حركة

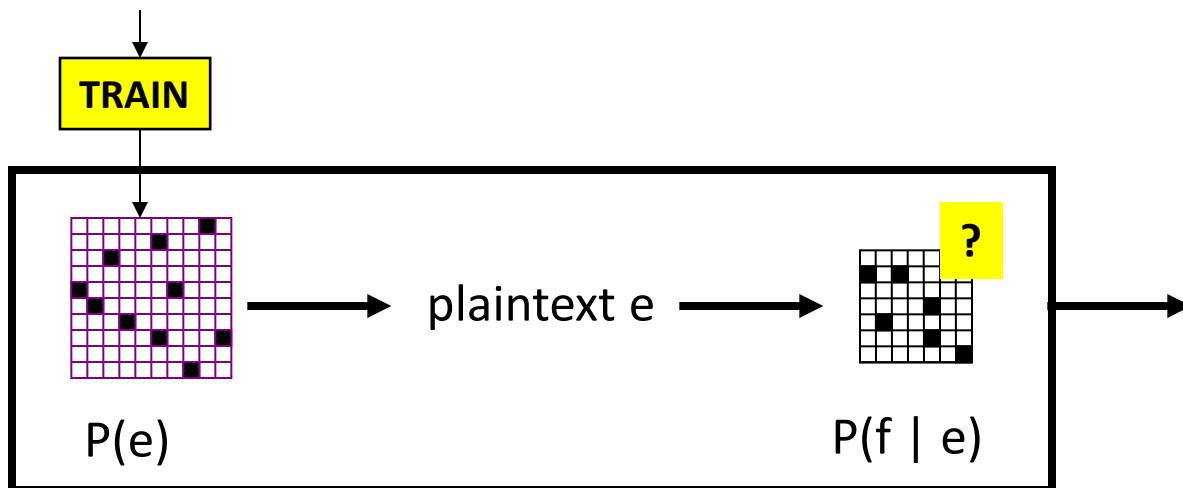
المقاومة الإسلامية حماس في الانتخابات التشريعية وقال عباس في مؤتمر صحفي على هامش مشاركته في القمة العربية-اللاتينية الأولى إنه يتمنى على إسرائيل احترام خيار الشعب الفلسطيني حتى لو فازت حماس بالانتخابات، وأضاف "إذا نجحت حماس أو فتح سيكون هذا خيار الشعب الفلسطيني، وعلى الجميع قبول هذا الخيار بكل ترحاب".

من جانبه شجب رئيس الحكومة الفلسطينية أحمد قريع الطابع الأحادي الجانبي للانسحاب الإسرائيلي من غزة، وأكد أن إسرائيل تريد مغادرة هذه الأرضي لتعزيز سيطرتها على الضفة الغربية.

وقال قريع في كلمة له خلال مؤتمر نظمته وزارة الأوقاف في رام الله "سينسحبون من غزة ولكننا لا نعرف ما هو شكل هذا الانسحاب وماذا سيتركون، وما هو مصير المعابر والحدود، وكل ذلك غامض لأنه قرار أحادي الجانبين".

Foreign Language as a Cipher

BAGHDAD, Iraq (CNN) -- Six bombings killed at least 54 Iraqis and wounded 96 others Wednesday, including 20 civilians who died as they lined up to join the Iraqi army in Hawija when a suicide bomber detonated explosives hidden under his clothing, Iraqi officials said. That attack in the town about 130 miles (209 kilometers) north of Baghdad also wounded 30 Iraqis, said Iraqi army Lt. Col. Khalil al-Zawbai. A car bombing in Saddam Hussein's ancestral homeland of Tikrit also killed 30 Iraqis and wounded another 40, Iraqi officials said. The Tikrit explosion



رفض رئيس السلطة الفلسطينية محمود عباس مجددا تصريحات وزير الخارجية الإسرائيلي سيلفان شالوم التي

قال فيها إنه يتمنى على إسرائيل إعادة النظر في انسحابها من غزة، المقرر أن يتم الصيف المقبل إذا فازت حركة المقاومة الإسلامية حماس في الانتخابات التشريعية وقال عباس في مؤتمر صحفي على هامش مشاركته في القمة العربية-اللاتينية الأولى إنه يتمنى على إسرائيل احترام خيار الشعب الفلسطيني حتى لو فازت حماس بالانتخابات، وأضاف "إذا نجحت حماس أو فتح سيكون هذا خيار الشعب الفلسطيني، وعلى الجميع قبول هذا الخيار بكل ترحاب".

من جانبه شجب رئيس الحكومة الفلسطينية أحمد قريع الطابع الأحادي الجانبي للانسحاب الإسرائيلي من غزة، وأكد أن إسرائيل تريد مغادرة هذه الأرضي لتعزيز سيطرتها على الضفة الغربية.

وقال قريع في كلمة له خلال مؤتمر نظمته وزارة الأوقاف في رام الله "سينسحبون من غزة ولكننا لا نعرف ما هو شكل هذا الانسحاب وماذا سيتركون، وما هو مصير المعابر والحدود، وكل ذلك غامض لأنه قرار أحدى الجان

Exploiting Giga-Scale Non-Parallel Text

Accuracy of learned
bilingual dictionary

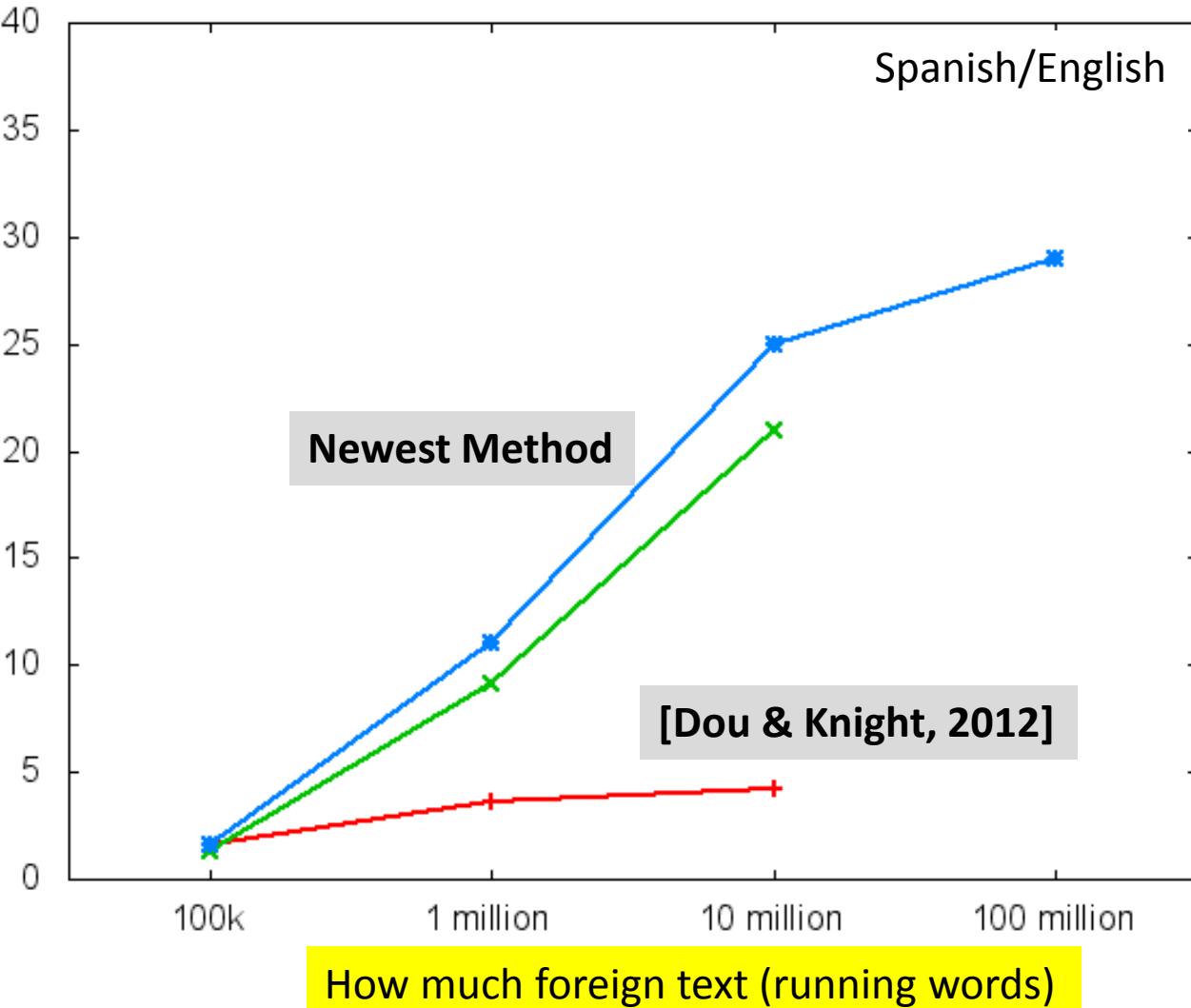
(Dou & Knight subm.)

not translations
of each other

English
text Foreign
text

deciphering
engine

bilingual
word-for-word
dictionary



Copiale Cipher

zmfzli[nx]̄[c]̄[n]̄[m]̄[c]̄[n]̄[b]̄[g]̄[r]̄[l]̄[k]̄[z]̄[ö]̄[ü]̄[v]̄[t]̄[n]̄[p]̄[z]
h̄[c]̄[j]̄[y]̄[s]̄[u]̄[t]̄[r]̄[l]̄[p]̄[i]̄[d]̄[f]̄[x]̄[d]̄[ä]̄[y]̄[l]̄[a]̄[g]̄[d]̄[r]̄[f]̄[i]̄[m]̄[z]̄[e]̄[c]̄[ü]̄[h]̄[r]̄[p]̄[i]̄[r]̄[d]̄[y]̄
ē[x]̄[k]̄[=]̄[r]̄[n]̄[ü]̄[r]̄[d]̄[c]̄[o]̄[z]̄[p]̄[l]̄[g]̄[d]̄[ö]̄[j]̄[n]̄[r]̄[h]̄[+]̄[b]̄[z]̄[ö]̄[s]̄[e]̄[h]̄[r]̄[m]̄[z]̄[=]̄[r]̄[ö]̄[m]̄[z]̄[/]̄[b]̄[n]̄[ü]̄[d]̄[;]̄[r]̄[j]̄[/]̄[s]̄[c]̄[o]̄[f]̄[l]̄[n]̄[t]̄[h]̄[l]̄[d]̄[o]̄[g]̄[r]̄[u]̄[p]̄[i]̄[h]̄[r]̄[e]̄[l]̄[e]̄[u]̄[s]̄[o]̄[d]̄[;]̄[z]̄[l]̄[b]̄[x]̄[l]̄[p]̄[i]̄[m]̄[z]̄[;]̄[ö]̄[l]̄[n]̄[g]̄[r]̄[p]̄[m]̄[t]̄[ö]̄[r]̄[u]̄[d]̄[;]̄[p]̄[s]̄[e]̄[n]̄[r]̄[u]̄[i]̄[t]̄[ü]̄[j]̄[d]̄[t]̄[i]̄[h]̄[c]̄[u]̄[g]̄[=]̄[r]̄[z]̄[g]̄[t]̄[a]̄[z]̄[l]̄[ü]̄[ä]̄[y]̄[g]̄[;]̄[w]̄[t]̄[q]̄[x]̄
j̄[i]̄[m]̄[=]̄[m]̄[p]̄[x]̄[c]̄[i]̄[h]̄[z]̄[c]̄[i]̄[x]̄[;]̄[i]̄[z]̄[g]̄[z]̄[n]̄[c]̄[ö]̄[l]̄[b]̄[n]̄[h]̄[+]̄[v]̄[z]̄[;]̄[é]̄[b]̄[g]̄[n]̄[p]̄[m]̄[z]̄[ü]̄[j]̄[z]̄[u]̄[h]̄[å]̄[l]̄[h]̄[=]̄[s]̄[w]̄[z]̄[é]̄[w]̄[;]̄[h]̄[é]̄[n]̄[m]̄[=]̄[b]̄[u]̄[d]̄[h]̄[l]̄[i]̄[n]̄[p]̄[b]̄[h]̄[z]̄[;]̄[i]̄[+]̄[p]̄[m]̄[i]̄[p]̄[+]̄[h]̄[r]̄[u]̄[i]̄[f]̄[ö]̄[l]̄[z]̄[n]̄[;]̄[n]̄[j]̄
ḡ[i]̄[n]̄[t]̄[v]̄[u]̄[z]̄[h]̄[z]̄[i]̄[z]̄[m]̄[;]̄[i]̄[+]̄[a]̄[r]̄[i]̄[c]̄[p]̄[i]̄[c]̄[;]̄[j]̄[r]̄[n]̄[q]̄[r]̄[f]̄[m]̄[p]̄[a]̄[w]̄[l]̄[p]̄[u]̄[n]̄[;]̄[p]̄[u]̄[;]̄[h]̄[i]̄[s]̄[=]̄[l]̄[o]̄[i]̄[r]̄[a]̄[j]̄[n]̄[r]̄[e]̄[m]̄[u]̄[b]̄[f]̄[h]̄[u]̄[w]̄[r]̄[p]̄[c]̄[e]̄[r]̄[l]̄[e]̄[z]̄[n]̄[;]̄[f]̄[c]̄[y]̄[é]̄[;]̄[l]̄[w]̄
z̄[ü]̄[h]̄[z]̄[i]̄[h]̄[b]̄[z]̄[y]̄[ä]̄[;]̄[u]̄[m]̄[h]̄[j]̄[p]̄[a]̄[l]̄[o]̄[s]̄[=]̄[g]̄[z]̄[+]̄[h]̄[g]̄[t]̄[ü]̄[h]̄[a]̄[l]̄[r]̄[l]̄[b]̄[h]̄[z]̄[j]̄[ö]̄[s]̄[+]̄[i]̄[;]̄[j]̄[z]̄[a]̄[t]̄[g]̄[z]̄[h]̄[m]̄[o]̄[d]̄[;]̄[x]̄[m]̄[p]̄[h]̄[é]̄[ü]̄[l]̄[v]̄[z]̄[ä]̄[;]̄[b]̄[c]̄[j]̄[ü]̄[p]̄[ä]̄[;]̄[f]̄[ä]̄[w]̄[ü]̄[r]̄[ö]̄[ç]̄[;]̄[u]̄[d]̄[ç]̄[z]̄[ö]̄[;]̄[l]̄[a]̄[z]̄[ü]̄[;]̄[n]̄[p]̄[ö]̄[d]̄[=]̄[l]̄[a]̄[z]̄[ü]̄[w]̄[x]̄[j]̄[ä]̄[;]̄[l]̄[v]̄[h]̄[g]̄[r]̄[z]̄[+]̄[b]̄[a]̄[r]̄[i]̄[n]̄[s]̄[t]̄[ä]̄[z]̄
m̄[ö]̄[r]̄[z]̄[w]̄[ö]̄[r]̄[p]̄[m]̄[t]̄[z]̄[ä]̄[h]̄[;]̄[A]̄[n]̄[i]̄[s]̄[u]̄[p]̄[;]̄[n]̄[;]̄[A]̄[d]̄[a]̄[t]̄[b]̄[;]̄[h]̄[c]̄[u]̄[p]̄[i]̄[r]̄[g]̄[z]̄[h]̄[e]̄
h̄[i]̄[=]̄[r]̄[j]̄[t]̄[ä]̄[b]̄[o]̄[;]̄[i]̄[r]̄[l]̄[ö]̄[l]̄[u]̄[w]̄[;]̄[h]̄[n]̄[é]̄[z]̄[b]̄[d]̄[ä]̄[;]̄[h]̄[i]̄[z]̄[o]̄[z]̄[k]̄[r]̄[;]̄[l]̄[p]̄[ö]̄[g]̄[g]̄[;]̄[u]̄[w]̄
p̄[j]̄[z]̄[ä]̄[r]̄[g]̄[a]̄[=]̄[g]̄[p]̄[v]̄[ü]̄[z]̄[c]̄[r]̄[g]̄[h]̄[c]̄[;]̄[ü]̄[r]̄[g]̄[h]̄[p]̄[m]̄[ö]̄[;]̄[u]̄[z]̄[u]̄[w]̄[r]̄[z]̄[f]̄[p]̄[ü]̄[r]̄[g]̄[z]̄
ḡ[p]̄[ö]̄[j]̄[u]̄[n]̄[;]̄[+]̄[i]̄[n]̄[u]̄[x]̄[;]̄[h]̄[i]̄[z]̄[n]̄[b]̄[s]̄[u]̄[n]̄[p]̄[r]̄[+]̄[i]̄[;]̄[a]̄[;]̄[d]̄[+]̄[t]̄[i]̄[m]̄[c]̄[f]̄[m]̄[j]̄[i]̄[z]̄[n]̄[
d̄[n]̄[ü]̄[;]̄[é]̄[l]̄[o]̄[s]̄[=]̄[l]̄[j]̄[y]̄[r]̄[;]̄[ü]̄[s]̄[e]̄[t]̄[p]̄[;]̄[d]̄[h]̄[i]̄[á]̄[z]̄[g]̄[f]̄[j]̄[m]̄[p]̄[c]̄[;]̄[;]̄[y]̄[d]̄[z]̄[a]̄[l]̄[o]̄[a]̄[;]̄[g]̄[â]̄[l]

zâl omi ntm fôj tçj pkr zrx dôc: i x m z h ôn A dôd ê fôd: uâ
é y è c: âi r m h i= p r ô y m p l i= p i h c s ân l o x i h i c l u b.
M p é o o + d x i / z c: u b.

M h i j k p c ân p l m t y é r g ô t c p r h i l i p r j y p l = m b z i z t a
u p r z: i z x z h b l d é y r é g c â h i c i r x ô b s z x p b u h i i p a x t p n p
w p e s h i c p r h c m l i d: i p e p r f j h i g z i p r l b t i y z s n p u d l a y i
h i n y d o i p m â z x p r o c i â p y ö u w d c f y z = p x u p u n f = z a s b x t
r p y é b m o i: u d h i m z t x h m i g h i p r u z a x g p w t h l a k d: i r u l p i
x u c u n x e i f h i r a l o d: p b = t u p u l a i d: j p m s i m p o o t h i c g i r
z a i b z e x u d h i s u m h e z x + e z y i l h i f m h i f r o d h i f h i l c u
f f l o b m u o r p e o o d c i h i c y x i a g x w z p i p u z h i z i o l p o i l a m
p e z x d l c i r l b q n r l g p y p p a z h d h c e l u d m z i j m p n x j c d a r f z
c z p f y o r u z u d t r i = l m p y a b d d a y x p m d z r m i d u b m i r p o
z k o c r a j m p c: z t h i c v e y r x u p i u l y p s x n | a c o o m d g i u z x +
w i n g r p + g p e = l c y p u d z h i z d p z | d r y y m m.

M p é o o + b d n.

C a p r o c e i p v o n r a r | A n â y x u l c e p r b g = z g h i z |

105 pages, 75000 letter tokens,
no word spacing, no illustrations.

Copiale Cipher

Section headers

Some scratch-outs, rare

h̄īc̄īj̄ȳs̄p̄ūt̄h̄l̄p̄r̄īḡd̄r̄j̄x̄b̄d̄ȳl̄āḡd̄r̄f̄+īl̄m̄z̄w̄c̄ūh̄īr̄p̄īr̄d̄ȳl̄
ē.l̄ī=p̄n̄ūr̄d̄c̄q̄l̄ḡd̄j̄n̄h̄+b̄z̄īōs̄s̄ēl̄p̄=r̄ōm̄z̄l̄l̄n̄d̄:h̄p̄ō
c̄p̄l̄r̄īh̄l̄d̄ḡr̄p̄ūs̄l̄r̄āl̄z̄ēl̄ūōd̄f̄:z̄āl̄b̄x̄l̄p̄īv̄ēf̄:ōl̄b̄ȳh̄m̄l̄ḡr̄
m̄p̄ōr̄ūōd̄f̄:s̄ēp̄ūīt̄ūj̄l̄t̄īh̄c̄ūḡ=r̄z̄ḡm̄āz̄ūl̄āȳḡ||w̄t̄q̄x̄
j̄m̄=m̄p̄x̄c̄īh̄c̄īr̄j̄īz̄ḡn̄c̄f̄l̄b̄h̄t̄+v̄z̄l̄|é̄b̄ōn̄j̄m̄z̄l̄j̄z̄ūh̄á̄l̄
h̄ī=f̄w̄z̄ēw̄|b̄c̄h̄m̄ȳl̄ūd̄h̄l̄īn̄p̄b̄h̄īr̄+p̄m̄r̄j̄+l̄p̄ūīf̄l̄z̄n̄||n̄j̄
ḡī+v̄ūp̄f̄z̄īz̄m̄||r̄+t̄ār̄īc̄p̄c̄:r̄n̄q̄||f̄m̄p̄āōl̄p̄ū+v̄z̄b̄r̄īj̄ū
p̄ū|l̄h̄s̄=l̄l̄ōr̄á̄j̄n̄īr̄m̄ūb̄s̄īh̄ūr̄c̄ēr̄l̄ēz̄n̄:f̄f̄c̄ȳé̄||w̄
z̄īh̄|l̄b̄z̄á̄|ūm̄h̄j̄p̄āl̄ō=ḡz̄q̄+l̄ḡt̄x̄h̄j̄āl̄h̄l̄||b̄d̄h̄j̄l̄ōx̄s̄ī:||z̄
ā+ḡz̄q̄ūm̄f̄||x̄m̄p̄h̄é̄ȳl̄n̄z̄á̄|b̄c̄īūp̄á̄||f̄f̄īūs̄p̄í̄|w̄ūr̄c̄ȳ||ū
d̄īx̄ō||l̄āz̄ū|v̄p̄ōd̄=l̄ūz̄āl̄ūx̄j̄ā||l̄v̄l̄ḡr̄z̄+b̄āl̄n̄āp̄á̄z̄
m̄r̄īz̄ō||ó̄p̄īm̄z̄q̄h̄||Δ̄n̄īs̄ūp̄||v̄|l̄d̄+b̄=h̄c̄ūp̄īr̄ḡz̄ē
h̄ī=l̄r̄īh̄t̄á̄b̄ōr̄:īl̄ōl̄ūp̄īr̄ēs̄b̄d̄j̄s̄h̄īz̄ōz̄h̄r̄:l̄p̄ōḡḡī=ūw̄
p̄īz̄á̄r̄ḡā=ḡp̄v̄īz̄c̄ēr̄z̄h̄c̄:ú̄r̄ḡh̄p̄m̄ó̄|ūz̄ūw̄r̄z̄f̄p̄ūīl̄=ḡz̄
ḡp̄ōj̄ūn̄||īn̄j̄īh̄āz̄b̄ūn̄p̄r̄+||ā|d̄+h̄īm̄c̄f̄n̄||īz̄n̄
d̄n̄ūm̄:é̄l̄ōs̄=l̄īj̄r̄|d̄ūēm̄f̄r̄:d̄h̄īá̄z̄ḡf̄r̄p̄īc̄:||v̄d̄z̄l̄ōd̄s̄:h̄

z̄á̄l̄ōm̄īr̄īm̄f̄j̄t̄c̄p̄k̄x̄r̄z̄k̄d̄ō:īx̄m̄z̄h̄ōn̄Δ̄d̄é̄f̄īā:ū
d̄é̄c̄:á̄īr̄n̄l̄ī=p̄ōȳm̄p̄b̄+t̄īc̄s̄á̄n̄l̄x̄z̄h̄īc̄l̄ūb̄.
Δ̄p̄á̄ō+d̄x̄īz̄c̄:ūb̄.
Δ̄h̄īz̄k̄p̄c̄á̄n̄f̄ām̄p̄é̄r̄ḡō+c̄p̄ūh̄īr̄j̄ȳr̄=m̄b̄z̄īz̄t̄
ūp̄x̄:z̄z̄k̄z̄h̄b̄d̄é̄r̄é̄ḡc̄á̄h̄īr̄x̄d̄l̄s̄z̄x̄h̄īs̄h̄īl̄īr̄p̄āx̄m̄p̄ūp̄
w̄t̄ēz̄l̄c̄p̄h̄ēm̄l̄ī:ūp̄ēp̄z̄f̄h̄ḡz̄īp̄r̄l̄b̄h̄īȳz̄z̄n̄ūp̄d̄āūp̄
h̄n̄v̄ȳd̄īl̄p̄m̄á̄x̄k̄p̄ōc̄īá̄p̄ȳd̄ūw̄h̄c̄f̄ȳz̄=ūx̄ūp̄ūn̄|=z̄l̄s̄b̄x̄
r̄p̄ȳé̄b̄īm̄ōī:ūd̄h̄īm̄z̄t̄x̄h̄m̄b̄h̄j̄p̄ūz̄āx̄ḡp̄w̄+r̄l̄x̄s̄l̄ī:z̄r̄ūl̄ī
x̄ūc̄ūn̄x̄ī||f̄h̄r̄á̄l̄ōh̄ī=p̄b̄+t̄p̄ūl̄īd̄:ȳp̄m̄s̄īm̄p̄ō+l̄h̄c̄ḡn̄r̄
z̄īr̄b̄z̄é̄x̄ūl̄h̄l̄s̄ūm̄h̄é̄x̄+ēz̄ȳīl̄h̄īl̄h̄t̄ōr̄d̄ūf̄h̄īl̄c̄ū
f̄l̄b̄m̄īr̄p̄ēō+d̄c̄īh̄īc̄ȳr̄x̄īāḡx̄w̄z̄r̄||n̄īz̄h̄īh̄īl̄p̄īl̄īn̄ī
p̄ēx̄d̄l̄īr̄l̄b̄q̄p̄l̄ḡm̄ȳp̄á̄z̄h̄īc̄é̄l̄ūf̄m̄z̄
c̄s̄p̄f̄j̄īr̄ūz̄||t̄p̄=l̄m̄p̄ȳá̄b̄d̄z̄ȳīp̄m̄z̄
x̄l̄c̄r̄á̄j̄m̄c̄:z̄+h̄īc̄v̄ȳr̄j̄īp̄īn̄d̄p̄s̄x̄n̄||
w̄īn̄r̄ī+ḡp̄é̄=l̄c̄ȳp̄d̄||z̄d̄f̄z̄||d̄r̄ȳm̄z̄||
Δ̄p̄á̄ō+b̄|l̄n̄.
C̄p̄r̄ōc̄é̄īp̄v̄ōr̄á̄||Δ̄n̄á̄ȳx̄ūl̄c̄ȳp̄b̄ḡ=z̄ḡh̄īz̄||

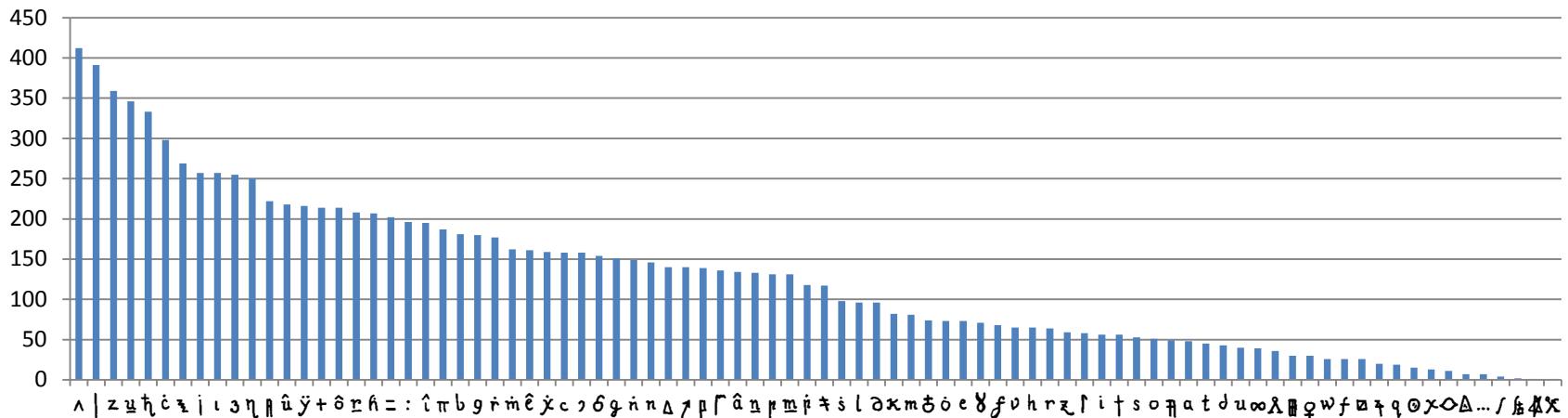
Preview text fragments
("catchwords")

Non-enciphered inscriptions:
Copiales 3 and Philipp 1866

Lines ≈
equal length

Paragraphs and section titles
always begin with
capitalized Roman letters.

Letter Frequencies



digraphs:

ን ከ 99

ኩ :

ኩ ለ 49

፡ ፩ 48

ዘ ዘ 44

trigraphs:

ን ከ አ 47

ኩ : ፩ 23

ኩ ዓ 22

የ ከ አ 18

ኩ ዓ | 17

tendencies:

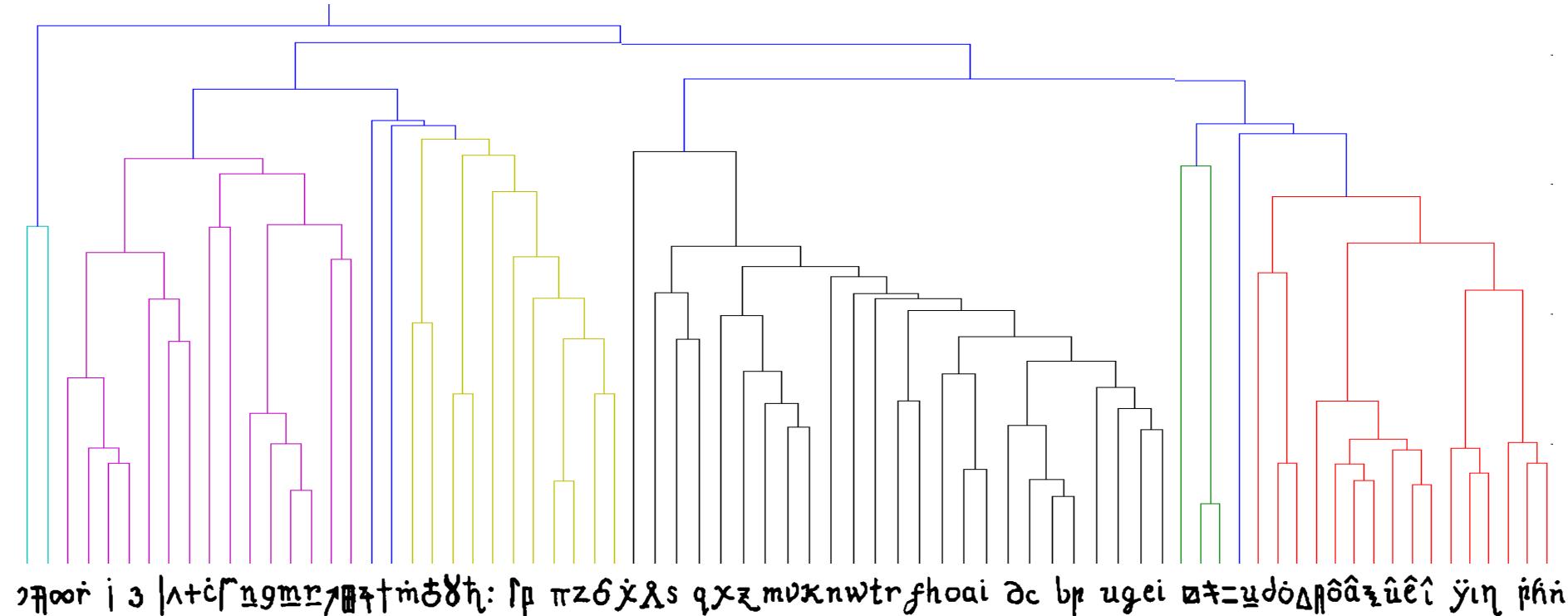
አ, ስ, ቤ, ሁ, ሪ followed by ዓ and ፍ

አ, ስ, ቤ, ሁ, ሪ preceded by ዓ and ፍ

Clustering of Cipher Letters

letters grouped if they have similar contexts (L/R neighbors)

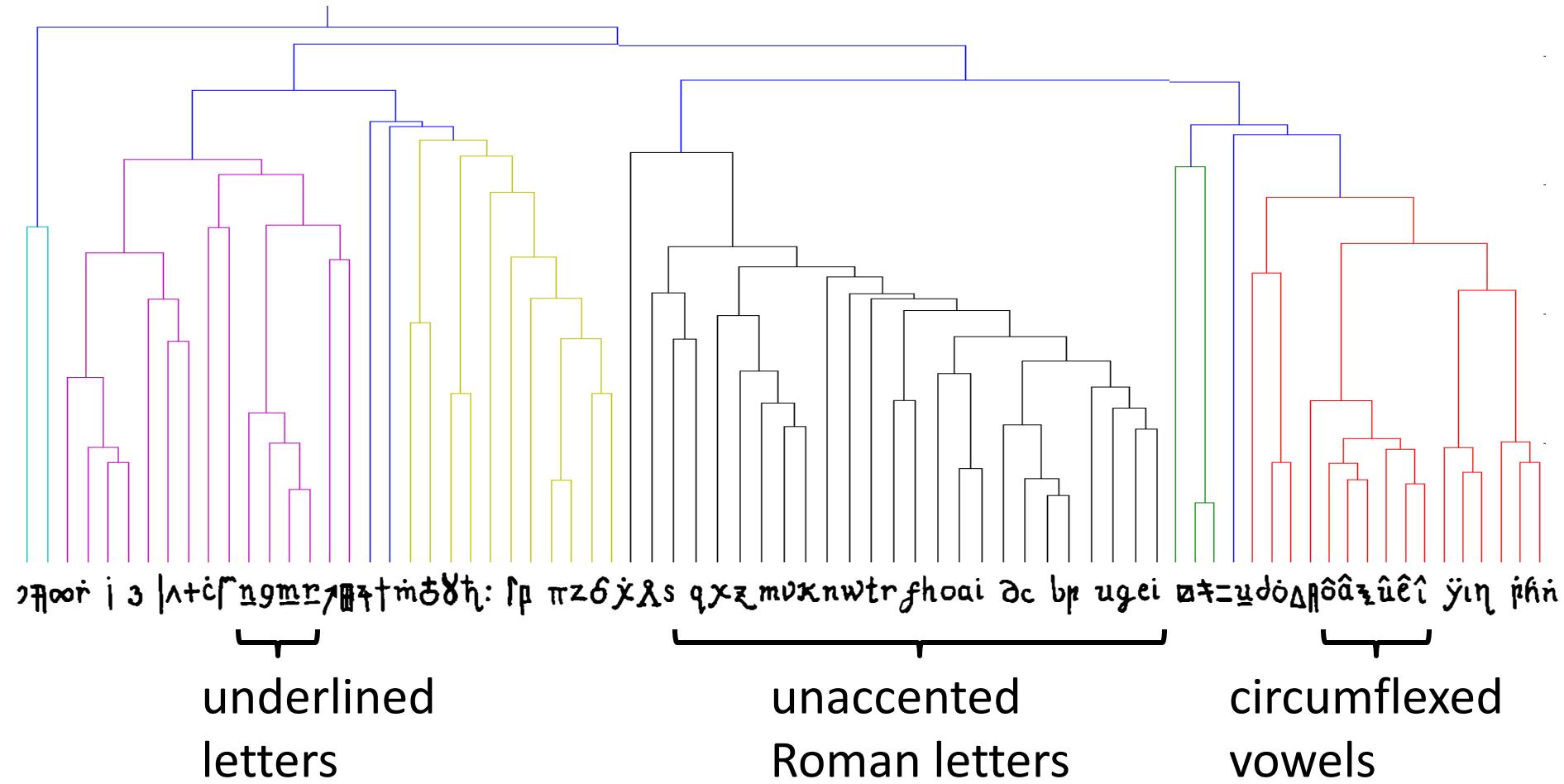
Scipy software



Clustering of Cipher Letters

letters grouped if they have similar contexts (L/R neighbors)

Scipy software



First Decipherment Approach

unaccented Roman letters that cluster:

a b c d e f g h i
k l m n o p q r s
t u v w x y z

most common letter = 12%
least common = very small

Kmûr:rzlôf|yj|hëi|h|z|l|n|p|â|z|b|A|g|z|=
i|z|l|z|z|p|c|l|â|r|g|K|l|h|p|r|h|p|l|z|j|n|p|û|d|r|h|l|a
=|g|z|w|p|y|ê|c|A|r|t|d+t|b|z|q|r|i|x|y|j|r|z|u|f|l|z|
p|z|j|p|d|r|i|f|z|u|s|z|p|l|d|m|z|q|g|â|K|h|=|l|x|o
o|r|i|z|l|b|z|l|f|u|m|y|j|z|v|z|â|x|j|x|p|l|p|l|z|h|l|c|t|o|g|g
z|u|t|h|z|f|u|x|e|z|g|h|l|h|p|i|h|p|l|i|z|t|h|f|r|o|y|m|â
+|h|h|r|z|o|z|n|b|s|q|t:z|r|K|p|d|R|h|l|c|u|l|g=|n|z|K
p|z|o|o|n|z|p|z|f|h|n|r|p|z|e|y|n|g=|n|p|g|t|d|z|n|z|p|K
p|n|j|i|x|y|r|i|g|p|u|x|l|b|g|l|i|t|b|d|n|f|j|h|z|o|l|e|z|n|o
f|u|r|c|f|z|f|n|l|b|n|h|t|m|d

xfgnlxknacbfzmk
lbuvcghtrhbkgzkn
fggnkbgbevb ...

Decipher against
80 plaintext languages.

Second Decipherment Approach

Homophonic cipher,

e.g.:

A = ፻ ፻ ፻ ፻ ፻ ፻

B = ሕ

C = ዕ ብ

D = ኃ

E = ድ ተ ፈ ት ቱ ቶ ፳

F = ም

G = ዷ



etc.

መስጠት፡ ከዚህ ዓይነት አረጋግጣት ለመተዳደሪያ መፈጸም አለበት
በመሸፍ የሚገኘው የአንቀጽ ስልጣን ጥሩ ነው፡ ይህንን የሚከተሉትን የሚከተሉትን መስፈርቶች
በመሸፍ የሚገኘው የአንቀጽ ስልጣን ጥሩ ነው፡ ይህንን የሚከተሉትን መስፈርቶች
በመሸፍ የሚገኘው የአንቀጽ ስልጣን ጥሩ ነው፡ ይህንን የሚከተሉትን መስፈርቶች

Homophonic Cipher

Result of computer attack on Copiale, using
80 possible plaintext languages?

FAIL

But, slight numerical preference for
German

Cipher Characteristics

digraphs:

, ī	99
č :	66
ī ^	49
: ü	48
z R	44

trigraphs:

, ī ^	47
č : ü	23
ī , ī	22
ÿ , ī	18
ī č	17

tendencies:

â, ê, î, ô, û followed by ɔ and ï

â, ê, î, ô, û preceded by z and ñ



should appear
adjacent in German text

Make full digraph table for cipher and for German

Key Observation #1

In Copiale, \mathfrak{C} almost always followed by \mathfrak{H}

In German, C almost always followed by H
(German CH is like English QU)

So guess: $\mathfrak{C} = C, \mathfrak{H} = H$

One Thing Leads to Another

ſt̄ = CH → ſt̄Λ = CHT → Λ = T ?

Each step is guesswork.

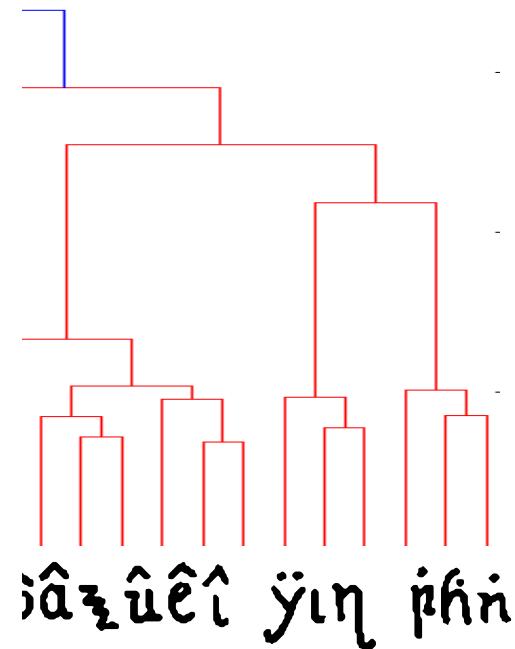
Must be willing to retract.

Weird task, not knowing German.

No longer care what the book says.

Cluster diagram crucial:

ÿ = | → ȳ = | , ȳ = |



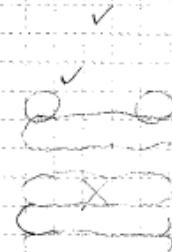
Spring Break 2011

c aeiou fpy dlmrztbvw hkngs j gmn ub

1
0

✓

? t h
i l ^
z T Z
s 8
g n r m
d = z
A o
w 3 j
c f
O m
P n h
e g u
c b p d
fallt by
z l P o f t w f h
z K n V M k
P f l ? x s e r fallt by vowels
z T S x



✓

✓

* der v cht z h n
und v e h i c : u
ein v s i h i c t
una ich y z n
cht v ich n z n
ich v che z h n
sch v t h m a
che v t c l a
ech v e i t + g n
die v s a h = p
rek ds z h m h
ine el t i n
gen t e l n
eit r e l i n
ver hen z h e l
ten lic z h e l
rei ant sch h m h
nte auf st i
ede ede i c h i
and v ein t h y e
den den die z h e l
run ter die z h e l
ter e i e r
sei h t e h i r : n
h t e h i r : n
hei n s g z i m a i c
n s g z i m a i c
ens i e t t ? n h
men h n t i d f m z
h n t i d f m z
h n t i d f m z
h n t i d f m z
ere e r e i c h
ere e r e i c h
das das h t m
rde nde h n k
ntie n t i e h e i
nge n g e s c h k z n
lte l t e s c h k z n
sch sch c s z
che che z v u
eie e i e t l a c
ede e d e t l a c

✓

✓

✓

?

(v)

vowels: u z H n i y
gmn e {nr} j
u {cb} a p i g R

Dö p n h

? z m c f + = z t A ? : s x z u f (w i g o f i d d l)
ao i a u f o
v w k s j u n o
need: f & y l m z u n o
(more in
german)
ut sp
at p
rf pe
(f)

ao i
v w
need: f &
ut sp
at p
rf pe
(f)

ge
k u
g u
g u
ng)

ru
3 i f
? 3 j c
? 3 j c
rm

Spring Break 2011

Cipher
letters,
in groups

German letters

c a e i o u f p y d l m r z t b v w h k n g s j g e r m a n u b

d	e	r	v	c	h	t	z	n	a	u
u	nd	v	e	h	i	c	ü	h	i	u
e	i	n	v	i	ch	ich	ü	z	ch	ü
u	na	g	h	ch	ich	ich	ü	z	ch	ü
g	cht	v	re	ch	che	che	ü	z	ch	ü
h	re	v	s	ch	sch	sch	ü	z	ch	ü
e	s	ch	v	ch	ech	ech	ü	z	ch	ü
ch	ch	ech	v	ch	ech	ech	ü	z	ch	ü
ü	die	v	die	ch	die	die	ü	z	ch	ü
z	re	ck	re	ch	re	re	ü	z	ch	ü
n	u	ck	re	ch	re	re	ü	z	ch	ü
ü	u	ne	u	ch	ne	ne	ü	z	ch	ü
ü	gen	u	ne	ch	gen	gen	ü	z	ch	ü
ü	ei	u	ne	ch	ei	ei	ü	z	ch	ü
ü	er	u	ne	ch	er	er	ü	z	ch	ü
ü	hen	u	ne	ch	hen	hen	ü	z	ch	ü
ü	lic	u	ne	ch	lic	lic	ü	z	ch	ü
ü	ten	u	ne	ch	ten	ten	ü	z	ch	ü
ü	rei	u	ne	ch	rei	rei	ü	z	ch	ü
ü	nte	u	ne	ch	nte	nte	ü	z	ch	ü
ü	an	u	ne	ch	an	an	ü	z	ch	ü
ü	af	u	ne	ch	af	af	ü	z	ch	ü
ü	ede	u	ne	ch	ede	ede	ü	z	ch	ü
ü	an	u	ne	ch	an	an	ü	z	ch	ü
ü	den	u	ne	ch	den	den	ü	z	ch	ü
ü	run	u	ne	ch	run	run	ü	z	ch	ü
ü	ter	u	ne	ch	ter	ter	ü	z	ch	ü
ü	er	u	ne	ch	er	er	ü	z	ch	ü
ü	sei	u	ne	ch	sei	sei	ü	z	ch	ü
ü	hté	u	ne	ch	hté	hté	ü	z	ch	ü
ü	hei	u	ne	ch	hei	hei	ü	z	ch	ü
ü	nsg	u	ne	ch	nsg	nsg	ü	z	ch	ü
ü	ens	u	ne	ch	ens	ens	ü	z	ch	ü
ü	men	u	ne	ch	men	men	ü	z	ch	ü
ü	hnt	u	ne	ch	hnt	hnt	ü	z	ch	ü
ü	ere	u	ne	ch	ere	ere	ü	z	ch	ü
ü	das	u	ne	ch	das	das	ü	z	ch	ü
ü	nde	u	ne	ch	nde	nde	ü	z	ch	ü
ü	nte	u	ne	ch	nte	nte	ü	z	ch	ü
ü	nge	u	ne	ch	nge	nge	ü	z	ch	ü
ü	te	u	ne	ch	te	te	ü	z	ch	ü
ü	sch	u	ne	ch	sch	sch	ü	z	ch	ü
ü	che	u	ne	ch	che	che	ü	z	ch	ü
ü	de	u	ne	ch	de	de	ü	z	ch	ü
ü	te	u	ne	ch	te	te	ü	z	ch	ü
ü	ac	u	ne	ch	ac	ac	ü	z	ch	ü

Grid

Spring Break 2011

Cipher
letters,
in groups

Quite a bit
of fooling
around →

German letters

c a e i o u f p y d l m r z t b v w h k n g s

German trigraphs

Cipher trigraphs

Grid

*	der	v	cht	ztn
	Und	v	e	:u
	ein	v	s	i
	una	v	h	z
	cht	v	ich	n
	reh	v	che	z
	sich	v	t	m
	che	v	i	c
	ech	v	it	g
*	die	v	sa	h
	rek	v	ds	u
	nde	v	er	h
	gen	v	t	a
	eit	v	r	o
	ver	v	ti	z
	hen	v	fe	h
	lic	v	le	e
	ten	v	ew	w
	rei	v	sch	z
	nte	v	st	m
	auf	v	an	h
	ede	v	ich	i
	and	v	ein	y
	den	v	nt	g
	run	v	ie	z
	ter	v	die	o
	fre	v	ze	u
	sei	v	rd	i
	hte	v	ri	n
	hei	v	nd	c
	nsg	v	ai	a
	ens	v	ki	l
	men	v	et	h
	hnt	v	et	z
	ere	v	ig	i
	ere	v	u	c
	das	v	tu	f
	nde	v	re	m
	nte	v	he	h
	nge	v	sch	z
	ite	v	ch	u
	ore	v	ce	z
	ede	v	de	v
	te	v	la	c

Trigraph
Decoding
Guesses

Spring Break 2011

Cipher
letters,
in groups

Quite a bit
of fooling
around →

German letters

c a e i o u f p y d l m r z t b v w h k n q s

German trigraphs

Cipher trigraphs

Grid

Partially deciphered text

?GEHEIMER?UNTERLIST?VOR?DIE?GESELLE
?ERDER?TITUL
?CEREMONIE?DER?AUFNAHME

Key Observation #2

unaccented Roman letters that cluster:

a b c d e f g h i
k l m n o p q r s
t u v w x y z

Kmûr:rzlôf|y, hêi hziln pâzba g z= iplz u kôc lârg k l h p r h p l z i n p u l d r h l a = g z w p y ê c A r t ð a + b z q r i x y j i r z u f l z p t i p d r i = f | u l s t p l d m | z n g â | k h = l h | l x ô o f : r i l b i f u t y j z v z â j x , r p l p i z h l c t ð o g g z û t p f p n x e z g h l h p i h p l i z t n f | r ô y m â + h h r z ô z n b s n t : z r k p d h h d c n l g = n z k p z o o n z p z f h n r p z e y n g = r p g t ð a z n z p k p n j i x y r i g p u l b g l i t b d n f p h h z ô l e z n ô f i n r c f z d n l b n h h m d

Actually, those are space bars

Initial translation:

First lawbook
of the **Φ** e **Θ**

Secret part.
First section
Secret teachings for apprentices.
First title.
Initiation rite.

If the safety of the **A** is guaranteed, and the **A** is opened by the chief **A**, by putting on his hat, the candidate is fetched from another room by the younger doorman and by the hand is led in and to the table of the chief **A**, who asks him:

First, if he desires to become **Φ**.

Secondly, if he submits to the rules of the **Θ** and without rebelliousness suffer through the time of apprenticeship.

Thirdly, be silent about the **A** of the **Θ** and furthermore be willing to offer himself to volunteer in the most committed way.

The candidate answers yes.

Voynich Manuscript (VMS)



- Medieval illustrated manuscript (early 1400s)
- 235 pages, 6 sections, 38k word tokens, 35 letter types
- Undeciphered



History of Voynich Manuscript



**William Newbold,
Polymath, PhD UPenn**



**Wilfrid Michael Voynich
book dealer**

- 1921 WV presents VMS + **Marci letter** mentioning Bacon, \$160k price
- 1921 Newbold & WV announce decipherment

One-Page Letter Tucked Into VMS

Reverend and Distinguished Sir; Father in Christ:

This book bequeathed to me by an intimate friend,
I destined for you, **my very dear Athanasius [Kircher]**,
as soon as it came into my possession, for I was
convinced that it could be read by no one except
yourself. The **former owner** of this book once asked
your opinion by letter Accept now this token
Dr Raphael, tutor in the Bohemian language to
Ferdinand III, then King of Bohemia, told me the said
book **had belonged to the Emperor Rudolf** and that
he presented the bearer who brought him the book
600 ducats. He believed the author was **Roger Bacon**,
the Englishman. On this point I suspend judgment
At the command of your reverence,

Joannes Marcus **Marci** of Cronland
Prague, 19 August, 1665(6?)



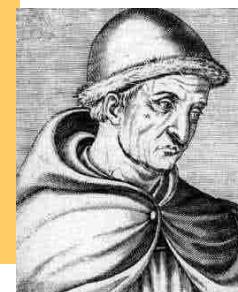
Kircher,
super-scholar,
recipient of
this letter



???,
owned VMS
before Marci



Emperor
Rudolf,
paid 600 ducats
for VMS



Roger Bacon
(1214-94)
“first scientist”

“I’m Not Francis Bacon”

History of Voynich Manuscript

1576-1612 Rudolf II purchases VMS

16xx Marci inherits VMS from ??

1665 Marci sends VMS to Kircher
with letter

1665-80 Kircher owns VMS

1680 Kircher dies

1921 WV presents VMS + Marci letter
mentioning Bacon, \$160k price
1921 Newbold & WV announce decipherment

History of Voynich Manuscript

1576-1612 Rudolf II purchases VMS

1608-1622 J. de Tepenecz signs VMS
in Bohemian court

**1630s George Baresch owns VMS
sends letter to Kircher**

1639 GB writes Kircher again

16xx Marci inherits VMS from GB

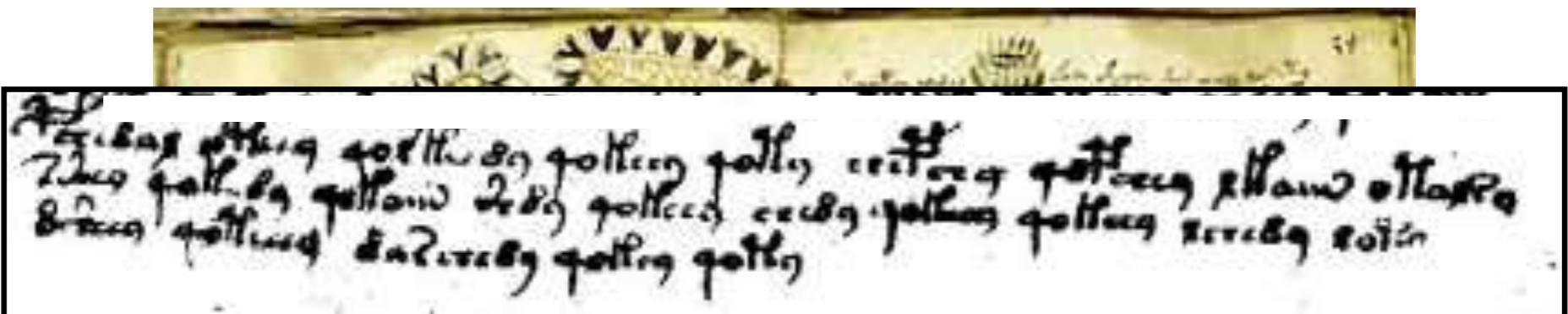
1665 Marci sends VMS to Kircher
with letter

1665-80 Kircher owns VMS

1680 Kircher dies

- 1864 Ethel Boole born in England
- 1865 WV born in Lithuania
- 1885 WV imprisoned, Polish nationalist
- 1890 WV & EB meet, marry in 1902
- 1898 WV publishes first book list
- 1912 WV acquires VMS in “ancient castle”
- 1914 WV moves to USA, opens bookshop
- 1919 WV sends photostatic copies of VMS
- 1919 Copying reveals de Tepenecz signature
- 1919 WV writes to Bohemian State Archvs
- 1921 WV presents VMS + Marci letter
mentioning Bacon, \$160k price
- 1921 Newbold & WV announce decipherment
- 1930 WV dies. VMS placed in vault, \$100k
- 1931 VMS appraised at \$19,400
- 1960 Ethel dies, VMS to secretary Ann Nill
“Castle” revealed as Villa Mondragone
- 1961 NY dealer Hans Kraus buys for \$24,500
- 1969 Kraus donates VMS to Yale
- 1972 Brumbaugh finds WV letters in BSA
- 200x Zandbergen finds 1639 Baresch letter
in newly online Kircher archive**

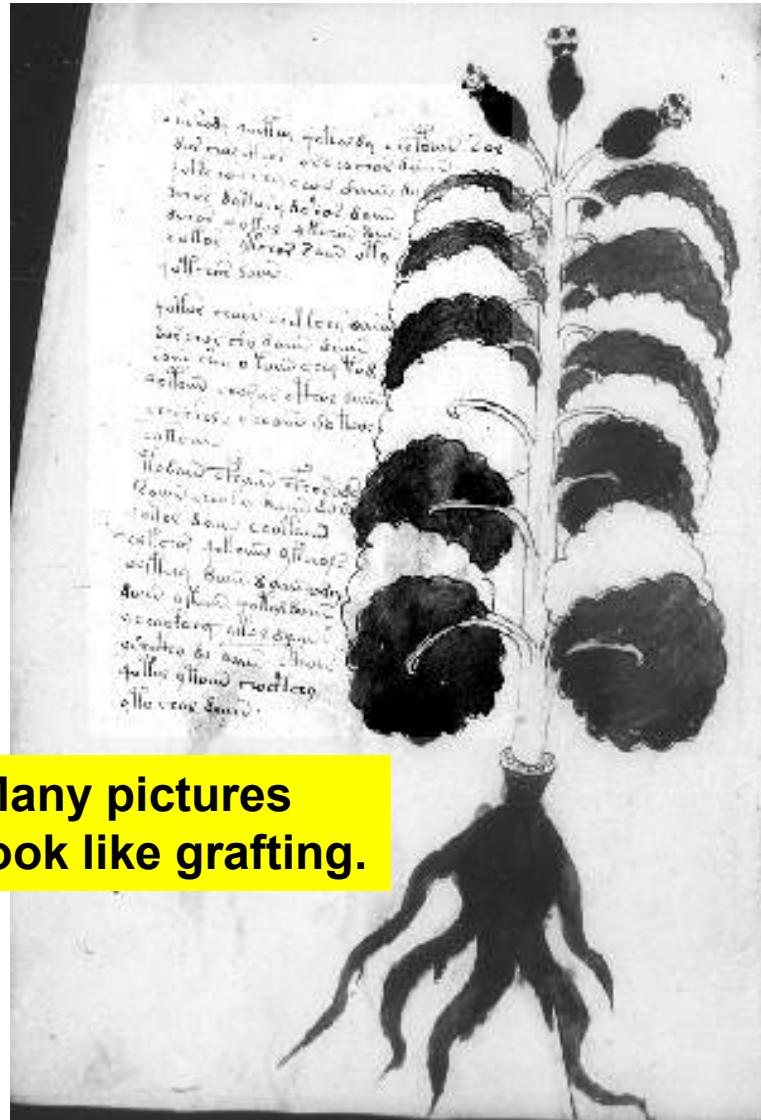
Voynich Manuscript (VMS)



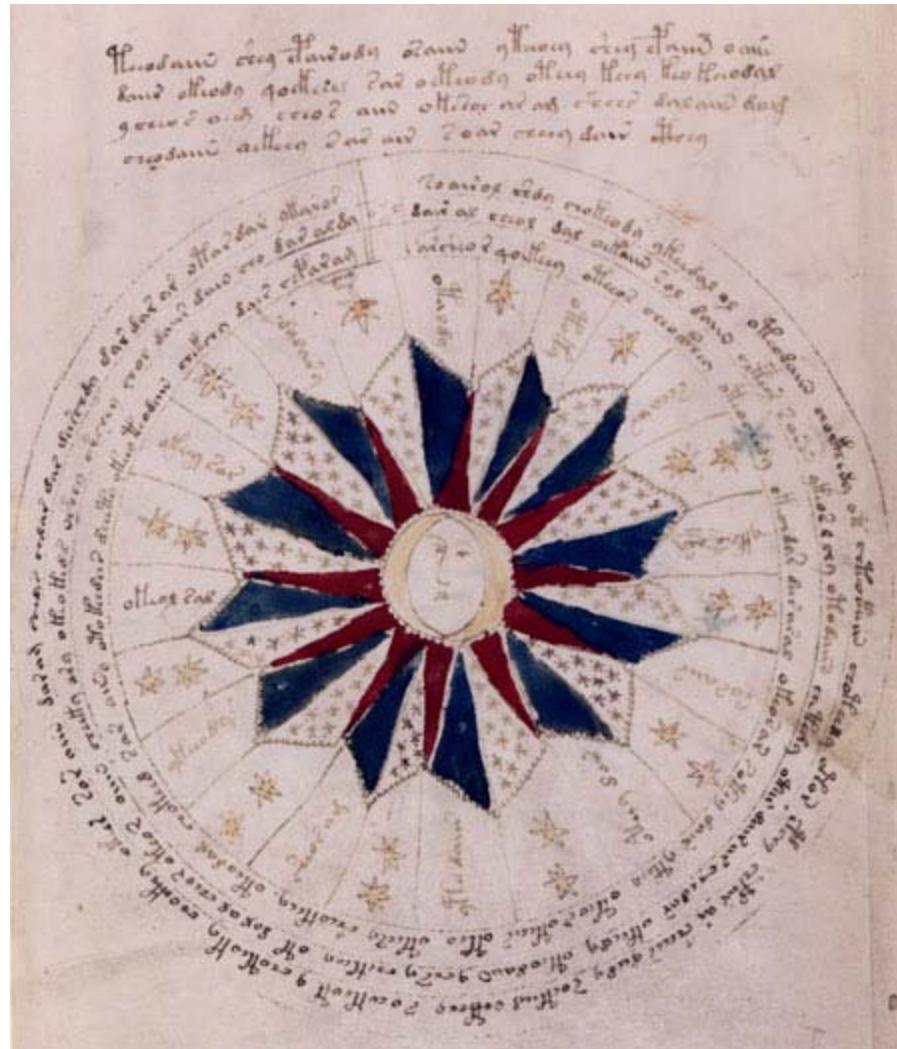
¶cc8a8 0¶cc9 408 ¶cc89 40¶cc9 40¶9 cc¶cc9 40¶cc9 2¶a111 0¶a829
2¶cc9 40¶cc89 40¶a111 289 40¶cc9 cc89 40¶cc9 40¶cc9 2cc89 20¶9
8cc9 40¶cc9 8a2cc89 40¶cc9 40¶9

BSC8AE OPCC9 4OE FCC89 40FCC9 4OP9 SCBS9 4OBSC9 EFAM OPAE29
2ZC9 4OFC89 4OFAM Z89 4OFCC9 SC89 4OFCC9 4OFCC9 ESC89 EOP9
8ZC9 4OPCCC9 8ARSC89 4OFC9 4OP9

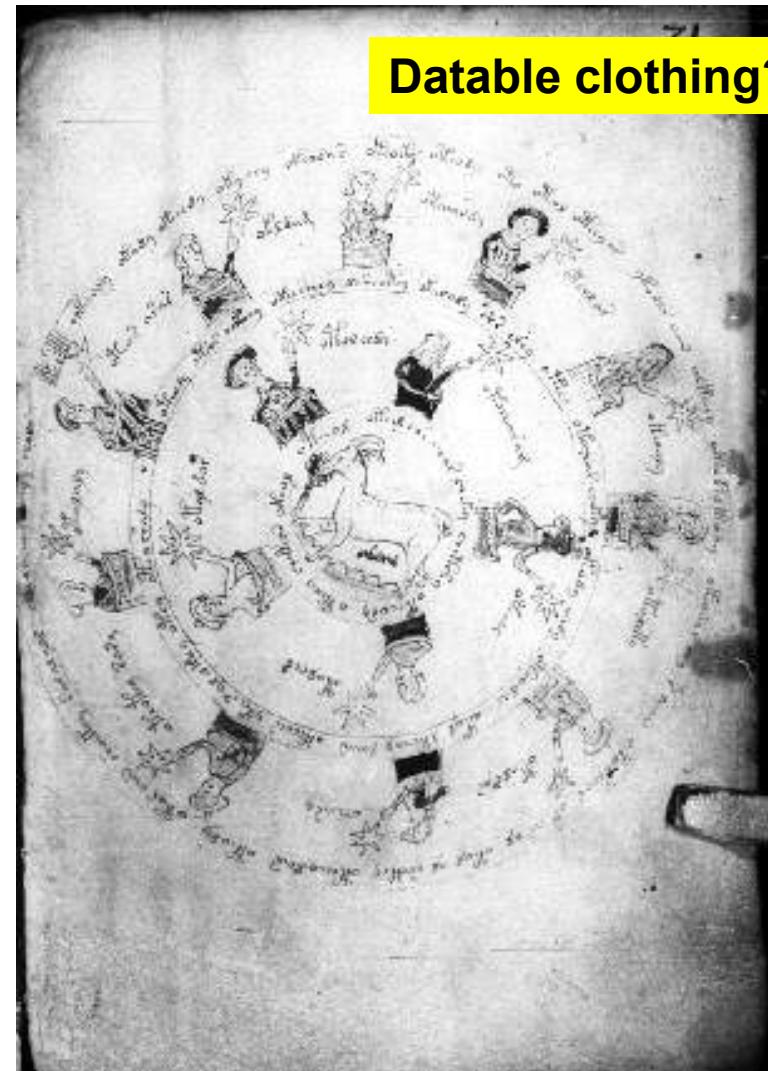
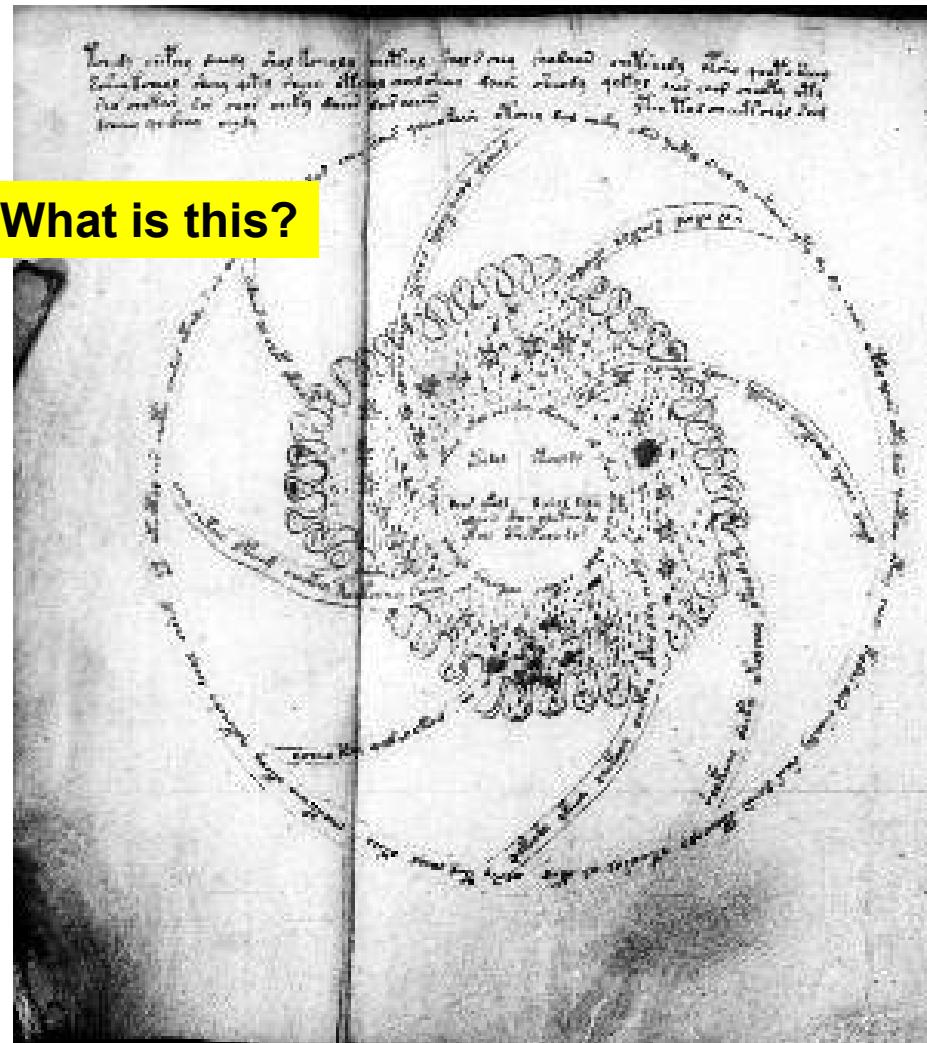
The Pictures: Herbal



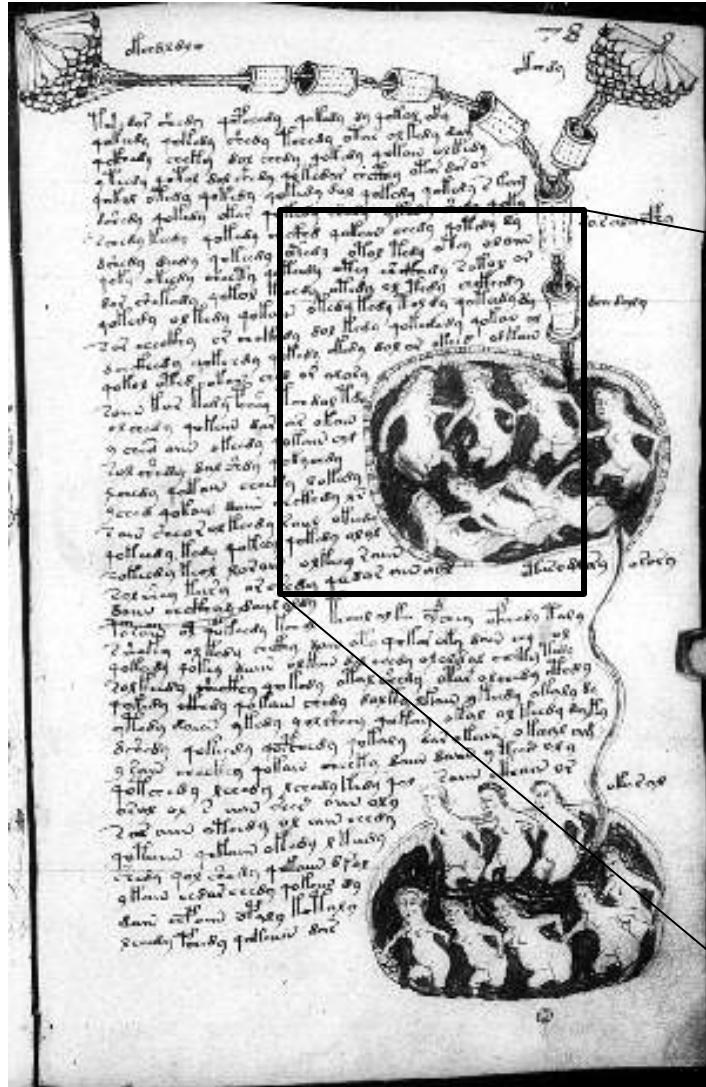
The Pictures: Astrological



The Pictures: Astrological



The Pictures: Biological



Small nudes in baths

Interconnecting tubes of liquids



The Pictures: pharmacological

medicine
jar?



The Text

- Approx. 38,000 words, unknown script
- Writing style similar to 15th century Florentine “humanist” hand
- Between 23 and 40 distinct characters
- No corrections, likely to have been copied
- Writing was done after illustrations

Transcription

ဘုရားရှင် ဂေါ်မြန်မာ ဂဲလီက ဂဲလီ၊ ခေါ်မြန်မာ ဂဲလီက ရှိမှုန်မြန်မာ
ဘုရားရှင် ဂေါ်မြန်မာ ဂဲလီက ခေါ်မြန်မာ ဂဲလီက ဂဲလီက ဂဲလီက ဂဲလီက
ဘုရားရှင် ဂေါ်မြန်မာ ဂဲလီက ဂဲလီက ဂဲလီက

၂၀၃၈၁၃ ၀၉။။၁၃ ၄၀၇ ၁၁၃၈၁၃ ၄၀၇။။၁၃ ၄၀၇။။၁၃ ၄၀၇။။၁၃ ၂၀၇၁၁၃ ၀၉၁၃၁၃
၂၀၇။။၁၃ ၄၀၇၁၁၃ ၂၀၇။။၁၃ ၄၀၇။။၁၃ ၄၀၇။။၁၃ ၄၀၇။။၁၃ ၄၀၇။။၁၃ ၂၀၇။။၁၃
၂၀၇။။၁၃ ၄၀၇။။၁၃ ၂၀၇။။၁၃ ၄၀၇။။၁၃ ၄၀၇။။၁၃ ၂၀၇။။၁၃ ၂၀၇။။၁၃

BSC8AE OPCC9 4OE FCC89 4OFCC9 4OP9 SCBS9 4OBSC9 EFAM OPAE29
2ZC9 4OFC89 4OFAM Z89 4OFCC9 SC89 4OFCC9 4OFCC9 ESC89 EOP9
8ZC9 4OPCCC9 8ARSC89 4OFC9 4OP9

Another medieval manuscript, just for calibration

A yonge man beryng flowers in his holmwolt ent
portmyd ⁸
Chasped for asynthe as yt ys a distannece betw
dynge to certayne pte of his zodylake as the denomy
nacion of hermyne þerwith vnyþythg other lareys
by an erer foundacyon of the denomynation þer he
maner / ffor in tyme .12 . signys of his zodylake ther
ben .4 . pte of a november many tymys satyned ~
rendryng the same or con signyfication . ffor tyme .12 .
signys saue þerst a synt part , þerut ys two signys
þerher makyngh a septyle aspecte and therfor ys so
callid by rans þt holsyng .6 . pte of his tyme
signys no .6 . Lymys goynge from com' rente they
cam' noted by seyt maner . * . whiche desposid in

Introduction to Astrology and Its Use in Weather Prediction, Medicine, and Agriculture, in English. Manuscript on Paper. 1490.

A yonge man beryng flowers in his holmwolt shal
portmyd ⁸

Chaspest ~~for~~ for a synghe as yt ys a distannece betw
dyngs to certayne pte of his zodiate as the denomyne
maner of hysme swyngis verywiche other lareys
by an eror fowdacyon of the denomyne maner yntys
maner / ffor in hi. 12 . syngis of his zodiate ther
ben . 4 . pte of a november many tymys satyned ~
rendryngs the same or con signifacions . ffor hi. 12 .
syngis satys first a synt parte , þerut ys two syngis
þerut makynge a seytys aspecte and therfor ys so
callid by rans ys holdyng hi. 6 . pte of hi. tyme
reignis no. 6 . Lyngs goynge from com' rente they
cam' noted by seyt maner . * . wchus disposed in

Alphabet: currier/d'Imperio

Transcription

c	π	շ
C	S	Z

Բ	Ֆ	Ց	Ֆ
P	F	B	V

Ք	Ճ	Ջ	Շ
Q	X	W	Y

յ	ա	չ	ր	օ	ւ	յ
J	A	E	R	O	I	D

Ճ	Յ	Ց	Շ	Գ	Ք	՞
6	7	8	9	4	2	

Կ	Ո	Ր
G	H	1

Ւ	Ո	Ր
T	U	0

Ո	Ո	Ր
N	M	3

Ո	Ո	Ր
K	L	5

Alphabet: currier/D'Imperio Transcription

c	π	շ
C	S	Z

Բ	Ֆ	Ց	Վ
P	F	B	V

Ք	Ճ	Ջ	Ւ
Q	X	W	Y

յ	ա	չ	ր	օ	ւ	յ
J	A	E	R	O	I	D

Ճ	Յ	Ց	Շ	Շ	Գ	՞
6	7	8	9	4	2	

Շ	Ռ	Ռ
G	H	I

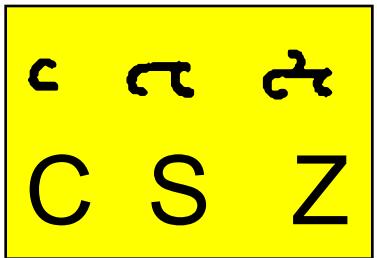
Ռ	Ռ	Ռ
T	U	O



Maybe this is really
IR IIR IIIR

There are several transcription
schemes to choose from.

Alphabet: currier/D'Imperio Transcription



Variations of ḷ, or separate characters?

᷑ π ḷ ḻ π ḻ ḻ

Alphabet: currier/d'Imperio

Transcription

c	ſ	ȝ
C	S	Z

ஃ	ஃ	ஃ	ஃ
P	F	B	V

ஃ	ஃ	ஃ	ஃ
Q	X	W	Y

Are these ligatures?

Is ȝ just a fancy way of writing ſ ஃ ?

If you didn't know English, how would you know if fi was the same as fi ?

Suppose fi **never** occurred. Would that be evidence?

Suppose fi did occur, with the **same** contexts as fi (e.g., *shing)?

Suppose fi did occur, but **never** in the same context as fi ?

Another common motif:

ତେବେରତ୍ତୋଳ୍ଯାଗର୍ଜ୍ୟ

Letter Frequencies

count	letter
25468	O
20227	C
17655	9
14281	A
12973	8
11008	S
10471	E
10026	F
6716	R
5994	P
5423	4
4501	Z
4076	M

25468	O
20227	C
17655	9
14281	A
12973	8
11008	S
10471	E
10026	F
6716	R
5994	P
5423	4
4501	Z
4076	M

count	letter
2886	?
1752	N
1413	B
1046	J
950	Q
908	X
591	T
524	*
431	V
316	I
217	W
157	D
156	3

2886	?
1752	N
1413	B
1046	J
950	Q
908	X
591	T
524	*
431	V
316	I
217	W
157	D
156	3

count	letter
148	U
96	6
74	Y
52	K
31	G
17	L
14	H
2	1
1	5
1	0

148	U
96	6
74	Y
52	K
31	G
17	L
14	H
2	1
1	5
1	0

Total
63k character tokens

most Frequent Words

count word

863	8AM
537	OE
501	SC89
469	AM
426	ZC89
396	SOE
363	OR
350	AR
344	SC9
318	8AR
308	4OFCC9
305	4OFCC89
283	ZC9
279	4OFAN
272	4OFC89
270	89
262	4OFAM
260	AE
253	8AE
243	2
219	SOR

count word

212	OFAM
211	8AN
191	4OFAE
186	ZOE
177	OFCC9
174	SCC9
172	SCOE
155	S9
155	OPC89
154	OPAM
152	4OFAR
151	9
151	4OE
150	S89
147	4OF9
144	ZCC9
144	OFAN
144	2AM
143	OPAE
141	OPAR
140	SX9

count word

140	OPCC9
138	OFAE
130	ZO
129	OFAR
119	ESC89
118	OFC89

etc

Totals:

8116 word types
38k word tokens

Word Length Distributions

Voynich

Length	Distribution
1	0.02
2	0.10
3	0.22
4	0.23
5	0.21
6	0.12
7	0.05
8	0.01
9	0.003
10	0.001
11	0.0001
12	0.00007
13	0.00002
35	0.00002

English

Length	Distribution
1	0.03
2	0.15
3	0.16
4	0.15
5	0.11
6	0.09
7	0.11
8	0.08
9	0.05
10	0.03
11	0.01
12	0.006
13	0.002

Counts on word types

Features of the Text

- 115 (out of 8116) word types appear doubled at least once

40lfcc89 40lfcc89

- 8 words appear tripled

40lfcc89 40lfcc89 40lfcc89

c0x c0x c0x

c2c0x c2c0x c2c0x

offaiv offaiv offaiv

ox ox ox

gHaiv gHaiv gHaiv

8aiv 8aiv 8aiv

40lfcc89 40lfcc89 40lfcc89

However, very few repeated word bigrams and word trigrams!

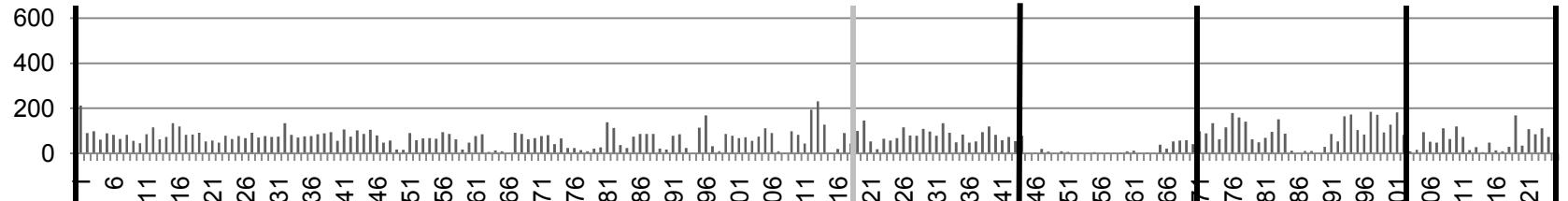
No word trigram appears more than 5 times.

Some Theories About the Text

- Cryptogram
- Phonetic writing system
- Philosophical language
- Outsider art
- Glossolalia
- Hoax

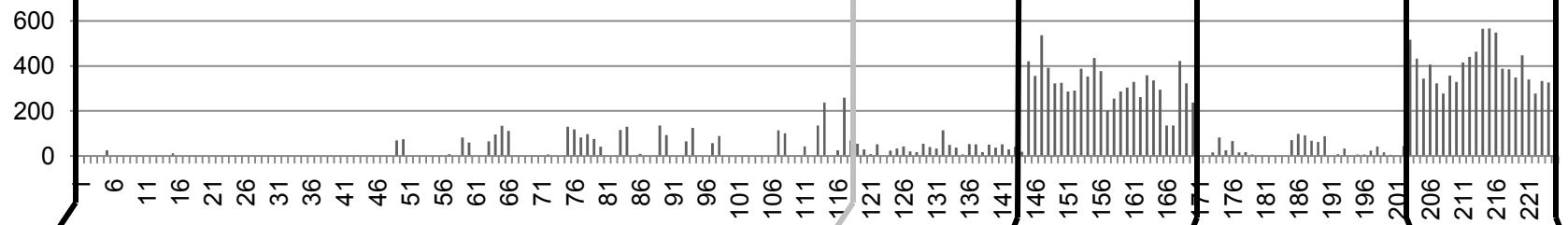
Generative models

Voynich words tagged as “a”



← pages →

Voynich words tagged as “b”



Herbal

Astro

Bio

Pharma

Stars

Known since Capt. Currier's analysis (1976):

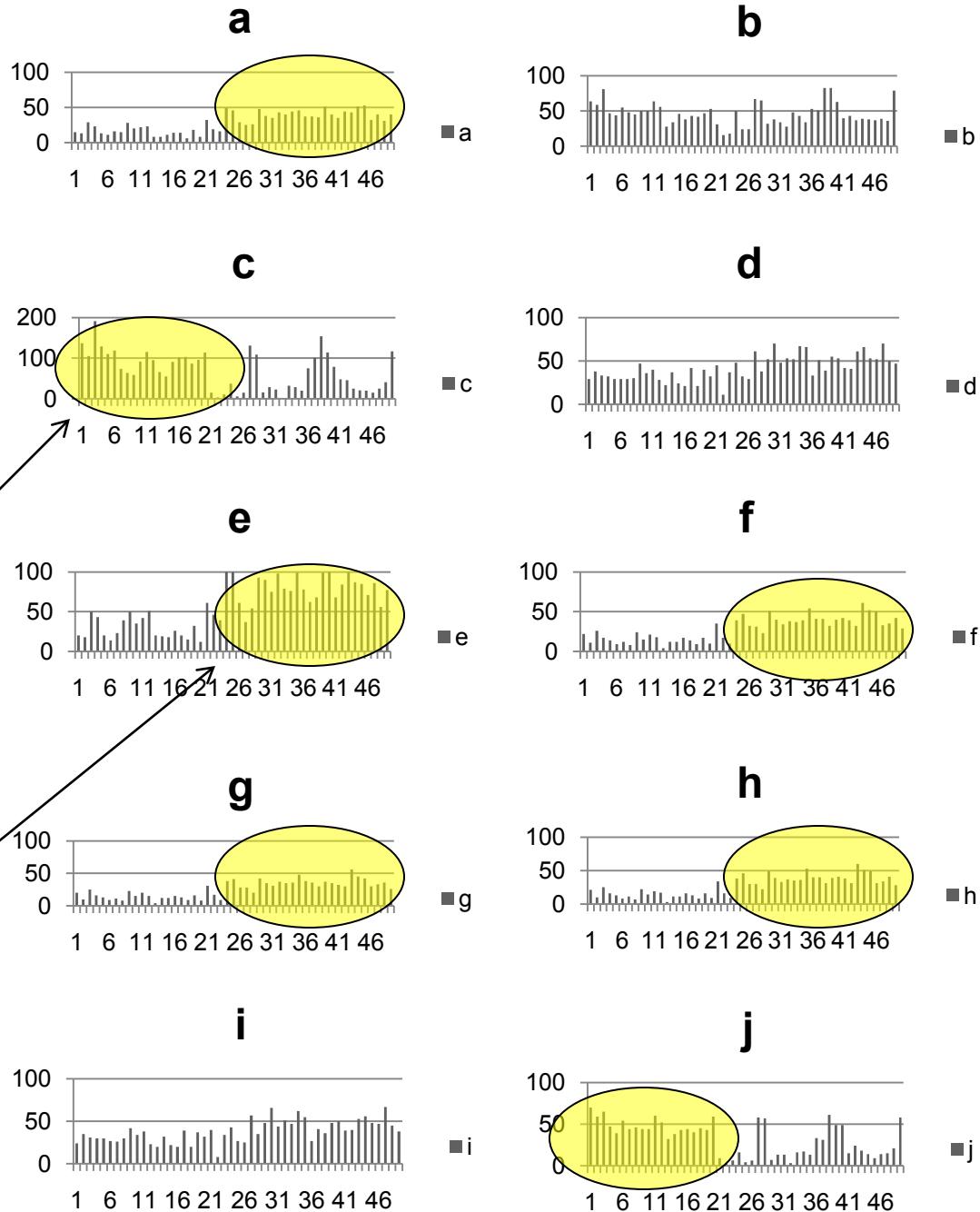
Two “languages” (in the formal sense).

Several handwriting styles, supposedly similar breakdown.

10 classes
of words:
Voynich-B

Tags per
page.

"Bio" words vs.
"stars" words



Substitution Cipher

Input	Best decipherment assuming plaintext is Spanish
cevzren cnegr qry vatravbfb uvqnytb qba dhvwbgr qr yn znapun ...	primera parte del ingenioso hidalgo don quijote de la mancha ...
VAS92 9FAE AR APAM ZOE ZOR9 QOR92 9 FOR ZOE89 ...	decos acho es imen des dena denal y des denta ...

If plaintext is assumed to be Latin:
quiss squm is onum pom
quuss hates s qum hatis ...

Conclusion

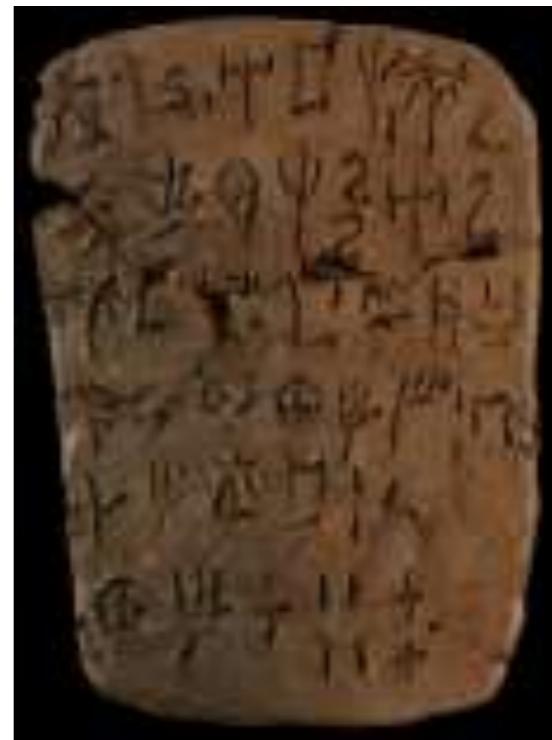
- Voynich Manuscript
 - What it is → pretty clear
 - Where it came from → less clear
 - What it means → totally unclear
- Lots of room for empirical, unsupervised computer techniques
 - Character analysis (e.g., ligatures)
 - Determining relations between words and pictures
 - Identification of “topics”
 - More cipher types

Undeciphered writing systems

Indus Valley
Script
(3300BC)



Linear A
(1900BC)



Rongorongo (1800s?)



Phaistos Disc (1700BC?)

