

Putting a Value on Comparable Corpora

Kevin Knight
USC Information Sciences Institute

joint work with

Beáta Megyesi, Christiane Schaefer (Uppsala)
Sujith Ravi (USC)

BUCC 2011

This Talk

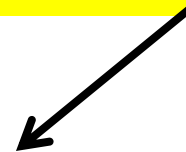
Our hero:
Warren Weaver



Statistical MT without
parallel text

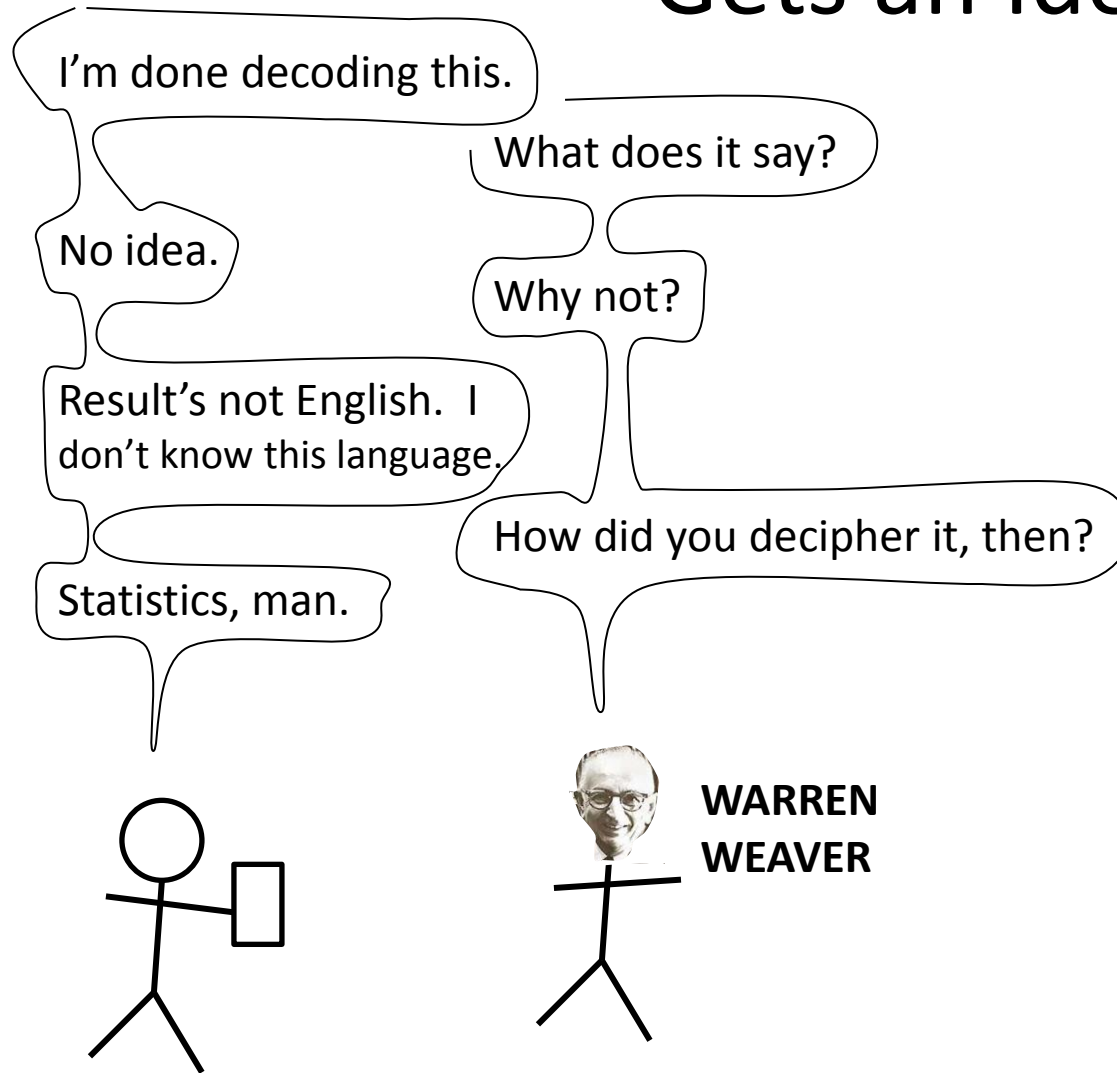
Comparable corpora &
code-breaking

Decipher this previously
undeciphered manuscript



=mzfzhi|nyx|cz||nêmlc||izxrbg|pax|zôÿl|ab+ηανπαζ
h|ε|ç|ÿ|Δ|u+ηα|π|ε|rgδp|zÿ|x|δ|á|ÿ|a|g|Δ|f+|amz|ε|c|û|η|r|p|ri|δ|ÿ|l
ε. h|=p|n|û|r|δ|ç|z||a|g|ô|j|v|η|η|+|b|z|l|ô|s|é|η|p|=r|ô|m|z||b|l|û|d:|η|p|σ
ç|ç||a|ri|η|n|d|g|ε|p|u|c|η|p|x|l|z|é|l|u|o|δ||:|ε|l|b|x|n|π|u|o|z||:|ô|l|b|ÿ|η|m|l|g|p|p

It's the 1940s, and Warren Weaver Gets an Idea



Weaver's friend only knew *statistics* of Turkish, not Turkish.

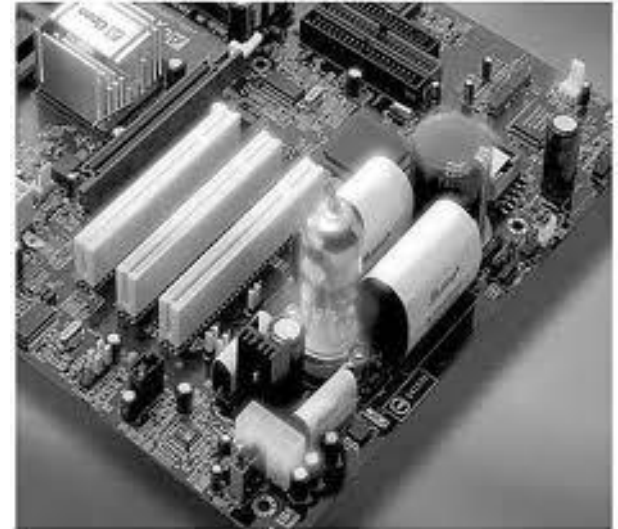
Weaver's friend soon replaced by a computer.

Could a computer *translate*, only knowing statistics?

A Note on Terminology



Pre-war computer



Post-war computer

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 ##### 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	7 #### V
Z	2 .

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 ##### 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	7 #### V
Z	2 .

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	7 ### 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	7 ### V
U	
V	
W	1 .
X	5
Y	7 #### V
Z	2 .

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

KDCY LQZKTLJKX CY MDBCYJQL: "TR

.o .a .e a . ee.e .o

HYD FKXC, FQ MKX RLQQIQ HYDL

a . . e .e .a

MKL DXCTW RDCDLQ JQMNKXTMB

. o.e a .a. o e.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 ##### 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 .

a o re.a r.a o o.e f

KDCY LQZKTLJKX CY MDBCYJQL: "TR

.o .a .e a freeze .o r

HYD FKXC, FQ MKX RLQQIQ HYDL

ar . f re .e .a

MKL DXCTW RDCDLQ JQMNKXTMB

. o.er a .a. o re.a r

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 .

a o re.a r.a o o.e f

KDCY LQZKTLJKX CY MDBCYJQL: "TR

.o .a .e a freeze .o r

HYD FKXC, FQ MKX RLQQIQ HYDL

ar . f re .e .a

MKL DXCTW RDCDLQ JQMNKXTMB

. o.er a .a. o re.a r

PTBMYEQL K FKH CY LQZKTL TC."

frequent cipher letters: ~~Q~~ ~~L~~ ~~K~~ C D T M ~~X~~ X

frequent English letters: ~~e~~ t ~~o~~ ~~a~~ n i ~~r~~ s h

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 .

a no re.air.a no no.e if
KDCY LQZKTLJKX CY MDBCYJQL: "TR
 .o .a n .e a freeze .o r
HYD FKXC, FQ MKX RLQQIQ HYDL
 ar ni. f n re .e .a i
MKL DXCTW RDCDLQ JQMNKXTMB
 .i o.er a .a. no re.air in
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 .

frequent cipher letters: ~~Q~~ ~~L~~ ~~K~~ C D T M ~~Y~~ X
 frequent English letters: ~~e~~ t ~~o~~ ~~a~~ n i ~~r~~ s h

a to re.air.a to to.e if
KDCY LQZKTLJKX CY MDBCYJQL: "TR
 .o .a t .e a freeze .o r
HYD FKXC, FQ MKX RLQQIQ HYDL
 ar ti. f t re .e .a i
MKL DXCTW RDCDLQ JQMNKXTMB
 .i o.er a .a. to re.air 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 .

frequent cipher letters: ~~Q~~ ~~L~~ ~~K~~ ~~C~~ D ~~T~~ M ~~X~~ X
 frequent English letters: ~~e~~ ~~t~~ ~~o~~ ~~a~~ n ~~i~~ ~~r~~ s h

a to repair.a to to.e if
KDCY LQZKTLJKX CY MDBCYJQL: "TR
 .o .a t .e a freeze .o r
HYD FKXC, FQ MKX RLQQIQ HYDL
 ar ti. f t re .e .a i
MKL DXCTW RDCDLQ JQMNKXTMB
 .i o.er a .a. 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 .

frequent cipher letters: ~~Q~~ ~~L~~ ~~K~~ ~~C~~ D ~~T~~ M ~~Y~~ X
 frequent English letters: ~~e~~ ~~t~~ ~~o~~ ~~a~~ n ~~i~~ ~~r~~ s h

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 .

Comparable, Non-Parallel Corpora

NOT PARALLEL

English corpus:

in the town where i was born ...

Enciphered

English message:

wpjk prtukuehbdg
diwokpd ...



English “w” =
Cipher “g”

NOT PARALLEL

English corpus:

in the town where i was born ...

Russian corpus:

ракета попала в
других ракет ...



English “missile” =
Russian “ракета”

Letter Based Ciphers

- Techniques undoubtedly worked out during 1950s-1960s at NSA and counterparts
 - Still classified!
- Active amateur community (e.g., ACA)
- Related to archaeological decipherment
- Some work in CL community:
 - [Knight & Yamada 1999, Knight et al 2006]
 - [Sproat, 2007]
 - [Ravi & Knight 2008, 2009abc, 2011]
 - [Sproat & Knight, 2009]
 - [Corlett & Penn 2010, Snyder et al 2010]
 - [Reddy & Knight, 2011]

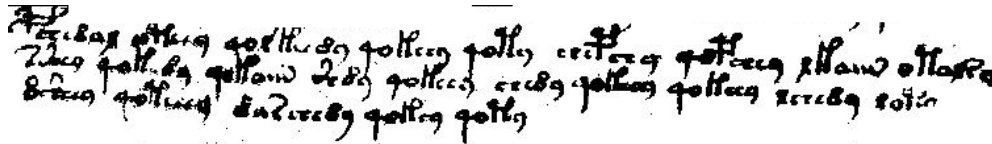
workshop

tutorial

ACL session

Some Unsolved Ciphers

- Voynich Manuscript (1400)



for more, see:
[Reddy & Knight,
ACL LaTeCH 2011]

- Zodiac 408 Serial Killer (1967)

Δ □ P / Z / U B □ X O R π 9 X π B
W V + 3 6 Y F 0 Δ H P □ K I D Y 3
M J Y Λ U I X Δ P T L N 0 Y D 0 0



for more, see:
[Ravi & Knight,
ACL 2011]

- FBI cipher (1999, released 2011)

ALPTE GLSE - SE ERTE
VLSE MTSE - CTSE - WSE - FRTSE
PURTSE ONPRSE WLD NCSE
WLD XCRCHSP NEWLD STS SE VI

- Kryptos (1990, new clue in 2011)

OBKR

UOXOGHULBSOLIFBBWFLRVQQPRNGKSSO
TWTQSJQSSEKZZWATJKLUDIAWINFBNYP
VTTMZFPKWGDKZXTJCDIGKUHUAUEKCAR

Copiale Cipher

=mzfzhi|ny+cz||nemlc+||izvrbgymk|sôyl+ab+qanpâz
 h|icjyân+u+qalp+rgdpyx|dâylagdpf+amz+zcûh|rprir|dyl
 e. h=anûr|dcoz||agôjvñh+bz|ôf|sêr||m=rômmz||bûd:qr|o
 c||ar|hndg+duc|qkx|ezêl+oδ||:abx|npiôz||:ôlby|hmlgr
 m|pôr|udô||:sêr|ui+ûj|d+n|hc|ug=rzgm|dab|lâyg||w+qx
 j||m=mp|ac|hôc|ir|x|izgn|cô||ab|h+vyz|é|bgn|mz|igz+ihâl
 h=|wzê|w. Abcm|dud|h|n|pb|h|r=+pm|r+th|ui|olzn||n||
 gn+vuz|hiz|m+|rê+ar|éc|ric:|r|nq|f|mp|áol|pû+vz|ôf|ju
 pu. Ah=|l|ô|râ|jn|rê|m|ub|f|h|v|rcê|p|lez|nn:f|c|yê||aw
 z|j|â. Abz|yâ|ûm|h|p|á|lo=|gzg+|igt|xû|h|l|qr|b|h|j|lô|x|s|t=|p
 a=+g|zq|môd||x|m|p|h|éy|vzâ|bc|j|u|pâ||f|q|s|u|o|p|t|w|û|rôc|y|u
 d|x|ô||Azû|v|pôd=|ar|d|u|v|x|jâ||Av|hgrz||b|d|rl|n|pâz
 m|r=|o|ô|p|m|d|mz|â|h||ni|u|p||v. Ad+bf|=h|cu|p|î|rgz|hê
 h=|r|r|h|â|b|h|r:|r|lô|u||q|h|é|z|b|ô|d|j|û|h|î|z|o|z|h|r:|p|ô|g|g|=|u|w
 p|j|zâ|rga=g|p|v|ñ|z|c|h|p|z|h|c:|û|rg|h|p|m|ô|j|u|z|u|w|p|z|p|û|r|l=gz
 g|p|ô|j|u|v|~~h~~||n|ñ|x|h|ø|z|h|ô|u|v|p|r|+||a|d|ô|h|m|c|t|m||i|z|n
 o|û|m:|é|l|o|s=|l|j|y|r|d|u|e|m|r|r:|d|h|â|z|g|q|r|m|p|c:||v|d|d|l|ô|d|c:|q

$$3\hat{a}^2$$

341 0 m i r m f e i + c h a x r z k d b c : x m z i n o n Δ d d e f c a : u d
 d y e c : a i r a n h i z p r o y m p b i r h e s a n a n x t h i c u b .

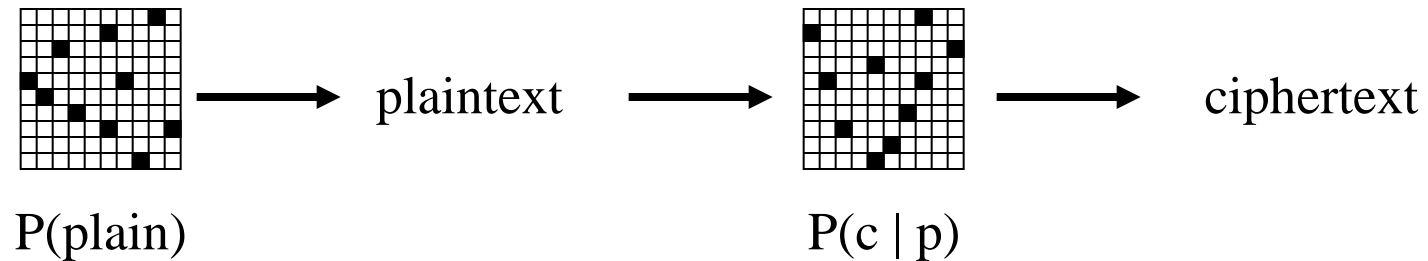
Ἡράκλεις + ὁ χί/ις: ὑβ.

[illegible]

Αφ' ου + b λη.

Сарвдѣи вѣрѣ и Анѣхъ и нѣхъ вѣрѣ = 39 хѣ

Computer Decipherment



Computer Decipherment

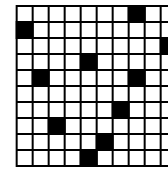
plaintext samples,
unrelated to ciphertext

LM TRAIN



$P(\text{plain})$

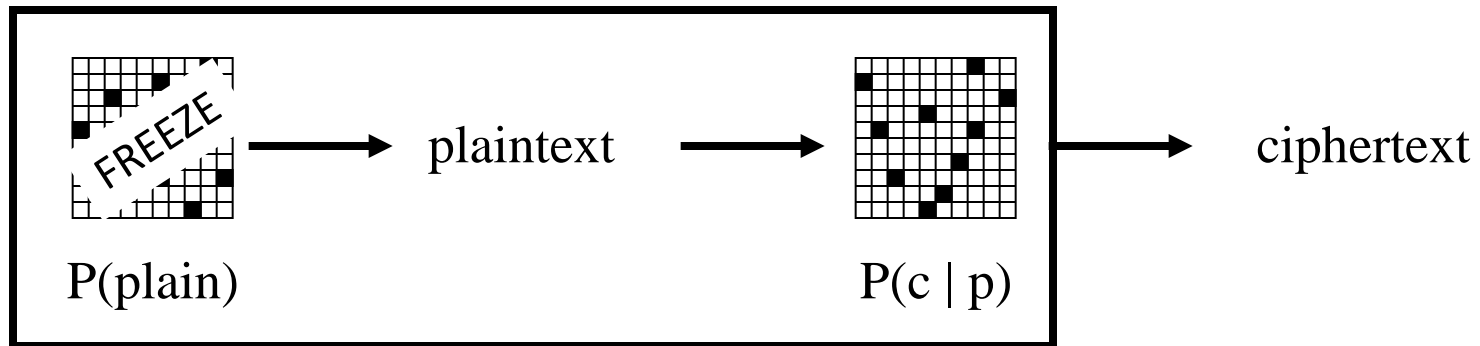
plaintext



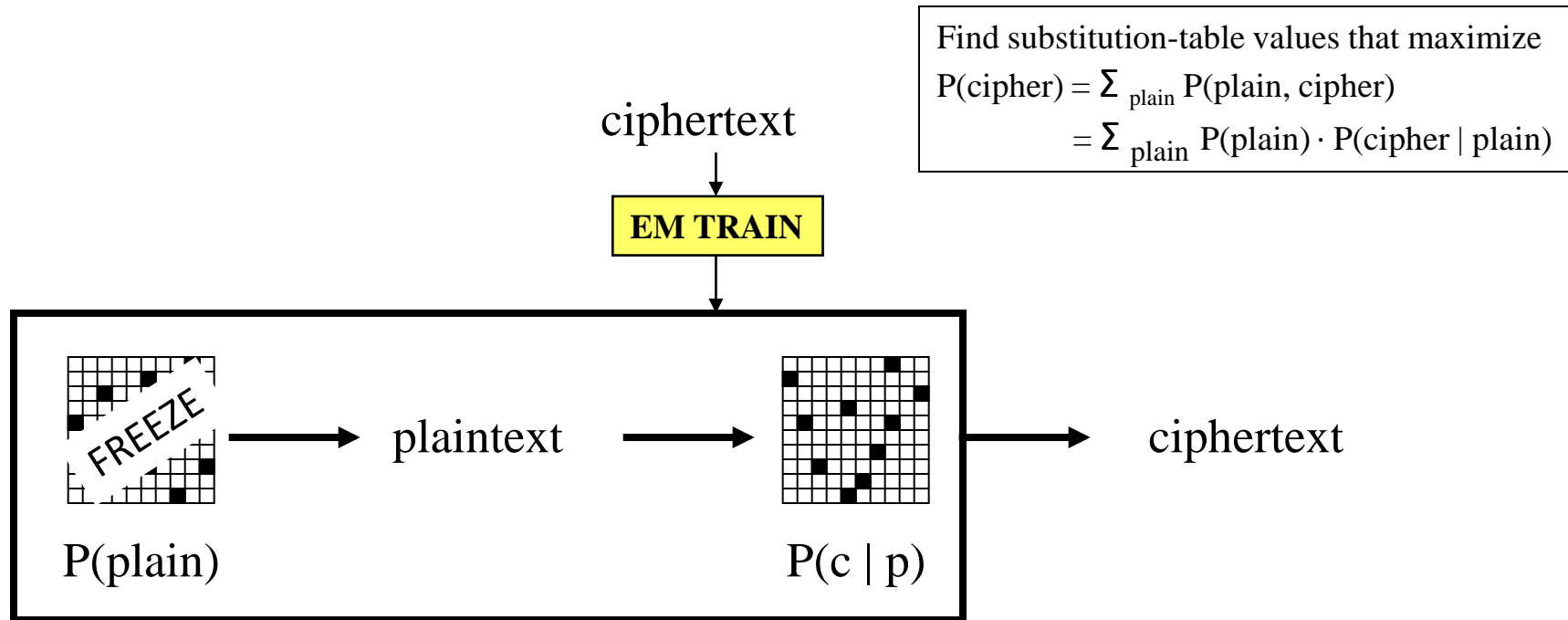
$P(c | p)$

ciphertext

Computer Decipherment



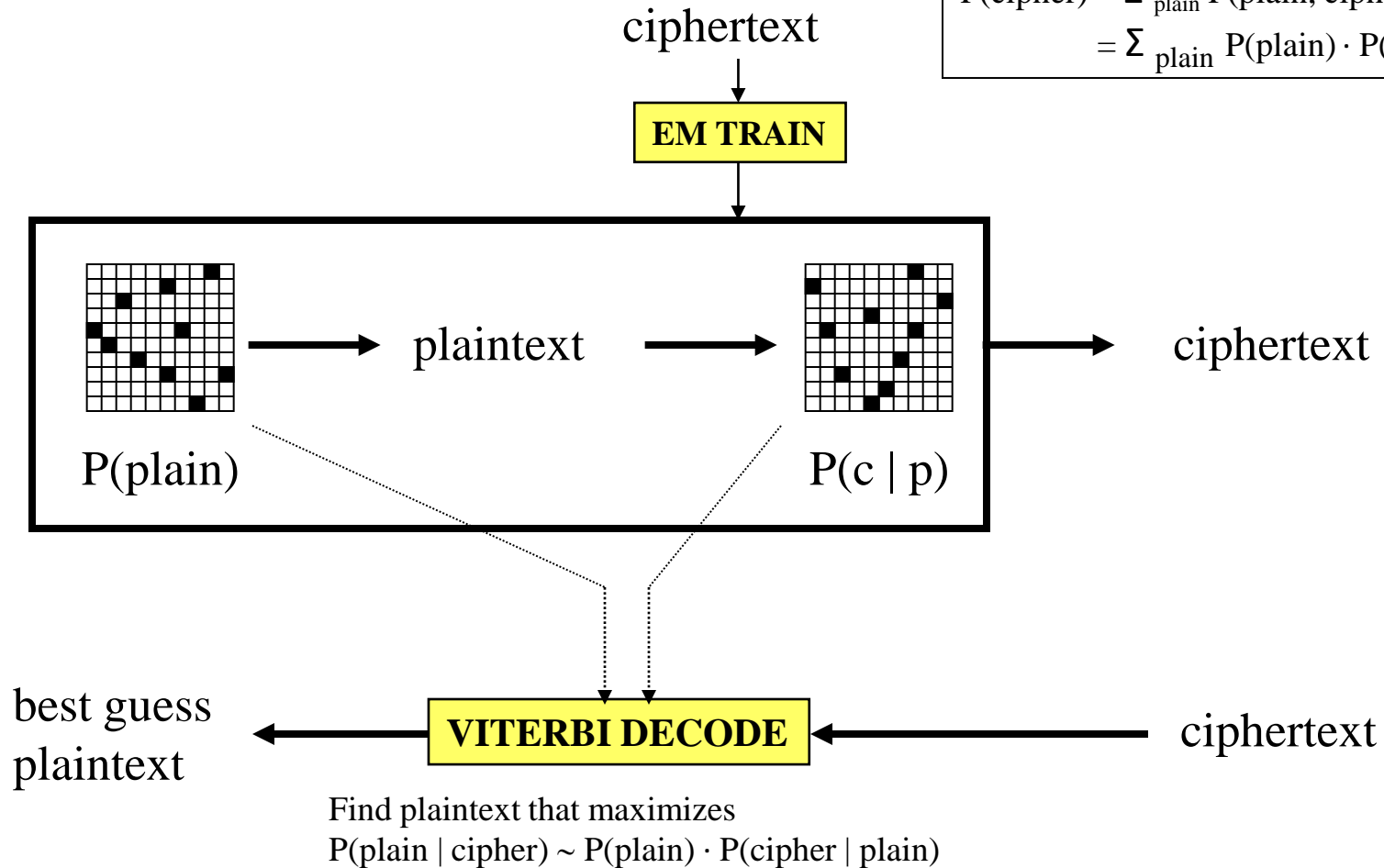
Computer Decipherment



Computer Decipherment

Find substitution-table values that maximize

$$P(\text{cipher}) = \sum_{\text{plain}} P(\text{plain}, \text{cipher})$$
$$= \sum_{\text{plain}} P(\text{plain}) \cdot P(\text{cipher} | \text{plain})$$



Copiale Cipher

=mzfzhi|nx|cz|nem|c|iz|rbg|ra|k|doy|ab+q|avpāz
 h|c|j|y|an|u|q|l|p|rg|dp|z|y|dā|y|lag|a|f|+|m|z|z|c|ū|h|r|p|ri|d|y|a
 e|f|h|=p|n|ū|d|c|p|z|p|g|ō|j|v|q|h|+|b|z|ō|s|ē|p|p|=r|ō|m|z|p|b|l|ū|d|:q|p|o
 c|p|l|p|h|p|d|g|p|u|c|q|p|a|k|z|ē|u|o|d|p|:z|b|x|p|u|b|z|p|:ō|l|y|h|m|g|r|p
 m|p|ō|r|u|d|ō|p|n|:s|ē|p|u|+|ū|j|d|+|n|h|c|u|g|=r|z|g|m|a|z|u|l|ā|y|g|p|w|+|q|x
 j|z|m|=m|p|x|c|h|z|c|r|x|i|=g|z|n|c|p|l|b|h|q|+|v|z|y|z|ē|b|g|r|m|z|i|g|z|u|h|ā|l
 h|=f|w|z|ē|w|A|b|c|h|m|u|d|h|h|n|p|h|z|r|=+|p|m|r|p|+|h|r|u|i|ō|l|z|h|n|p|j
 g|h|+|v|u|z|p|h|z|z|m|u|r|ē|+|a|r|ē|c|p|c|:p|r|n|q|r|p|m|p|ā|o|l|p|ū|+|v|z|ō|f|j|u
 p|u|A|h|s|=l|ō|r|ā|j|n|r|ē|m|u|b|f|h|u|v|r|c|ē|r|p|a|e|z|n|n|f|f|c|y|ē|f|l|w
 z|i|j|ē|A|b|z|y|ā|ū|m|h|j|r|ā|i|l|o|=g|z|g|+|h|g|t|x|ū|h|l|q|r|l|b|h|z|j|l|ō|x|s|+|:p|z
 a|=+|g|z|h|m|ō|p|p|x|m|p|h|ē|y|l|v|z|ā|b|c|r|u|m|p|ā|l|f|z|f|u|o|p|i|w|ū|r|ō|ē|y|u
 d|x|z|ō|A|z|ū|v|p|ō|d|=l|q|z|l|u|v|x|j|d|A|v|h|g|r|z|p|+|b|d|r|l|p|a|p|ā|z
 m|r|ē|o|ō|p|m|a|m|z|l|ā|h|A|n|i|u|p|z|v|A|d|+|b|f|=h|c|u|p|i|r|g|z|h|ē
 h|=r|r|h|t|ā|b|ō|r|r|:r|l|ō|l|u|z|q|h|ē|z|b|d|j|ū|h|z|ō|z|h|r|:l|p|ō|g|g|=u|w
 p|j|z|ā|r|g|a|=g|p|v|ū|z|c|h|p|z|h|c|:ū|r|g|h|r|m|ō|l|u|z|u|w|r|z|p|ū|i|l|=g|z
 g|p|ō|j|u|v|~~h|z|z|p|~~h|z|z|p|n|z|x|p|h|z|p|z|h|d|u|v|p|r|+|z|o|d|+|h|m|c|f|m|p|i|z|p
 d|h|ū|m|:ē|l|o|s|=l|j|y|r|u|e|m|p|r|:d|h|i|ā|z|g|q|r|m|r|c|:p|p|d|z|l|ō|l|ā|c|:q

zāl

zāl o m i r m f i j + c h r a x r z k d o c : i x m z i r o n Δ d o d e f c a : u a
 d y e c : a i r m n h i z p r o y m p l + r i h c s a n a n x i h i c l u b .

Αρ ά ο ο + d x i : i c : u b .

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

Αρ έ ο ο + b A n .

Ca p n d o c e i r v o n r a r A Δ n a y x u l c n p b g = z g h i x i

Simultaneous ID and Decipherment

- Cipher:

FZFZ VMJVLH DV RPMOVG VMOZM JZDZULZM KPV
DV VMNLAUH NHG OHIH NZLAGH VM NPZGOH DV
IANV MP SABH KPV DZ KPAVLV

- What language is this a code for?
- What is the decoding?

Method: For various plaintext languages, run EM to get

$$\begin{aligned} P(\text{cipher}) &= \sum_{\text{plain}} P(\text{plain}, \text{cipher}) \\ &= \sum_{\text{plain}} P(\text{plain}) \cdot P(\text{cipher} \mid \text{plain}) \end{aligned}$$

↑ different $P(\text{plain})$ for each plaintext language

Simultaneous ID and Decipherment

cipher: FZFFZ VMJVLH DV RPMOVG VMOZM JZDZULZM KPV DV VMNLAUH NHG OHIH NZLAGH VM
NPZGOH DV IANV MP SABH KPV DZ KPAVLV

Plaintext Language	P(cipher)	1-best Decoding
English	135.6	oron thicas of behity thind inormand bet of thecums any ings prarys th pervis of pual he full bet or beucat
French	133.5	meme angart la jussan ances gelibres que la aydroit fon cont perint as duenct la vide nu hoit que le quiere
German	130.8	jede arbeit zu vorden urder begewiem von zu urlicht wen stat weicht ur poenst zu alle po scht von zu volein
Italian	131.2	dive allaro ha sultad altel libegrel qua ha esprito con como perito al puento ha fica su divo qua di supare
Portuguese	122.9	dada esfazo de juster estas famancas que de esprino por todo parilo es pualho de live ou fino que da quiaca
Spanish	117.9	momo espera de justen estos podotros que de escriza fan tala cogina es cuonta de libe su vida que ho cuiera
Latin	134.2	vivi habera de suatem estis fidignis que de exprogi fam haci ficomi es puilla de cone su coní que di quoere

correct: mama espero le gusten estas palabras que le escribo con todo carino es
cuanto le dice su hijo que la quiere

Copiale Cipher

[illegible]
$$Z^{\wedge}$$

341 0 m i r m f e i + c h a x r z k d b c : x m z i n o n Δ d d e f c a : u d
 d y e c : a i r a n h i z r o y m f b i r h e s a n a n x t h i c u b .

Ἡράκλεις + ὁ χίλις + ὁ β.

Αηι ι κ ρ α η ρ μ π γ ε ρ θ ο + ι η ρ ρ η ι η ρ ι γ ρ λ = μ ο β ζ ι σ τ α
 υ ρ ι α : ι ε κ ζ η ι β δ ε γ ρ ε γ α η ρ ι ρ χ ο β σ : χ ρ ο υ η ι ρ π α χ μ π ν η ρ
 μ π ε ζ ι ρ ι ρ θ ε μ λ ι δ : η ρ ε π ρ ε ρ η γ ζ ι π ρ λ μ ι γ ζ ν ρ ι δ | α υ ρ ο
 η ν γ γ ο δ ι η μ α ε ρ ρ ο ι α ρ γ ο υ ω δ ρ γ ζ = η χ α π υ ν ι = ε λ σ β α χ
 ρ π γ ε β μ ο ι : μ ο η μ ζ ι χ η μ η γ η ρ ι ρ ζ α χ γ π ω η ρ λ α δ ι : ε ρ υ λ ρ ι
 χ υ ε υ η χ ι λ η ι ρ α λ ο χ ι ρ β = ν π υ λ ι δ : γ ρ μ σ μ ρ ο ο + η ι η ε γ ρ ι ρ
 ζ ι ρ β ζ ε χ υ | ο η λ σ υ μ η ε ε χ + ε ζ γ υ λ λ ι : μ η ι η ι ο ρ ο υ η η η ι λ α υ
 ρ ι ο β μ ι ο ρ ρ ε ο ο + δ ε ι η ε γ ρ ι χ ι α γ ε ζ ω ρ ι η ν ε ζ ι ο λ μ ι ο ρ ι α π
 π ε ζ α δ ι ρ ι μ β γ η ρ λ γ μ γ ρ α ζ η ρ η ε λ υ | α μ ζ ι η μ ρ η χ ι ο δ α ρ ζ
 ε ρ ο γ ο ρ ι υ ζ υ . Δ τ ρ ι : μ π γ α β δ ε ζ ι χ ρ μ ο ε ρ μ ι ο | υ β μ ι ρ ι ο ζ
 α | ο ε ρ α ι μ ρ ι ε : + η ι η ι ν ε γ ρ χ υ π ι ν δ ρ ρ α η | α ο μ δ γ ι υ ζ κ +
 ω ι η ρ ι + γ ρ ε = λ ε γ ρ μ ο ζ η ζ δ ρ ι ο ρ γ μ μ ι .

$$A \neq \infty + b \cdot \ln.$$

Σαρκοειρήνην ἄρ. Ἀναΰχινος ἔργ. = 39 h 17

Copiale Cipher

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

Section headers

Lines ≈
equal length

Paragraphs and section titles
always begin with
capitalized Roman letters.

Non-enciphered inscriptions:
Copiales 3 and Philipp 1866

Some scratch-outs, rare

Preview text fragments
("catchwords")

hí:cíjyánu+ηλπiγδρζχδδγλαγαρf+ιmzεcúhírríndéyá
ε. h=πuírδcφzηgδóvηh+bzδsέηp=ρόmzηbλúd:ηρ|σ
cφ||λrηhδgεμcηpκλzέλuδδ||:ελbηπiδb||:δλbγhímgíρ
mπóρuδδ||η:σέηp+újδ+ñhεuq=rzgmδzλuáýg||w+qχ
||m=mpacíhεcyrχ||εgzncφ||λbηh+vzy||έbgnírmzίgδzιhál
h=íwzέ|w. Λbcηmδuδhíhñπbηr=+pmpíρ+híρu|δolzn||ηqj
gn+vuz||hízim+re+ariέcric:πrñqρsmράoolπú+vzδbίju
πu. Λhs=|δi-râjnrέpμb|fhuνrcέipaeznñ:f|cýé||λw
zίjñ. Λbzýá|úmhíρáιo=gzg+liqzúhíh|ηr|bδh:íλδxst+:||z
a=+gzηímod||xmpíhεýλvzá|bcíumπál||qz|uotí|wúrócy|u
δxδδ||Λazú|vπóδ=ληzλvχjδ||Λvñgrz||+bδrλnδπáz
mór:íoolóρmδmzηáhΔn|fupz|v. Λδδ+bf=hcúπírgzηέ
h=írríh+ábδr:íρδλu|ηhεzδδj|úhíozhí:lpó|ggz=uw
pizárga=gπvñzénrχhí:úrgíhpmδ|uзуwpzπpúrl=gz
gποfuv~~híh~~ñηχhíhεzηδvπr+ι|α|δδ+himc+fm|fízn
δhúm:έλoδ=λjýr|δuemíρ:δhíázgρf|mpic:||vδzλoδac:η
zál

zálomírmíhí+chípxrzkδoc:íxmzηónΔδδέfca|:úδ
δýéc:áipnli=próym|b+ñhεsáηλhχhíhclub.
Hράoo+δxí|ic:ub.
Αhí|xpcáηpamπýêrgó+chíρúhíhíρiýpλ=mbbzístz
upx|:zεxzhí|bδéýrégcáhírcíxδb|s=χ|δuñ|irπαxmpv||p
wpéδ|úεpηδémλíd:ηρεπ|zfhígzιπrλbmýzñúpúδ|auíδ
ηvýgδ|ηmáaxpóciápyδuwbδcýz=ηxupuv|=zλsbx+
rπýébmoí|:uδhñmz|fχhíhíhíρuzaxgπw+ηλxδf:íru|pí
xúcuñxi|fñrálδh=rb=+vπuλíd:ýpmsmíoo+híhcgñr
zírzbzέx|δhλsúmhéx+ezý|hí=ímhíh+óróúhíhíhclú
f|obmúδrπέoo+δc|hεcýr|íagxwzr||ñú+δz|δlπóδrδm
pézxd|fíρλbgníρgmýpázhíhcléu|amz
czpδýδrúzu. Λtrí=ímpýábδaxj|pδv
x|ócá|mpic:χ+híhεvέyry|píñδpδxñ|
wíhír||gπé=acýpúδ|zδp|δrýpam.
Hράoo+b. Δn.
Cαρνδcέipvδhíρár. ΔAnáyχúncíhíρbg=zghíx|

Transcription Scheme

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99	â ch é oh ê eh ↑ ih ô oh û uh	î c. ï h. ï. n. ï. n. ï. s. ï. x. ï. y.	: . bar ... 3 three ∞ inf	∂ del Δ tri γ gam ι iot ∧ lam π pi	Δ tri.. ∅ lip ∞ nee ⊙ o.. ✱ ster ✕ bigx
25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99			↗ arr 7 bas 4 car + plus † cross ♀ fem ♂ mal Ⓕ ft Ⓖ no P sqp 222 ✂ pipe ✓ longS	H or H grr H grl H grc ↑ hk H sqi	
				Δ tri. (p8)	

Plus capital letters.

"Philipp 1866" - spelling limited to

πôjvhdtriarr
ehthreec.ahnarr
lamuhb lipuuro..
zs

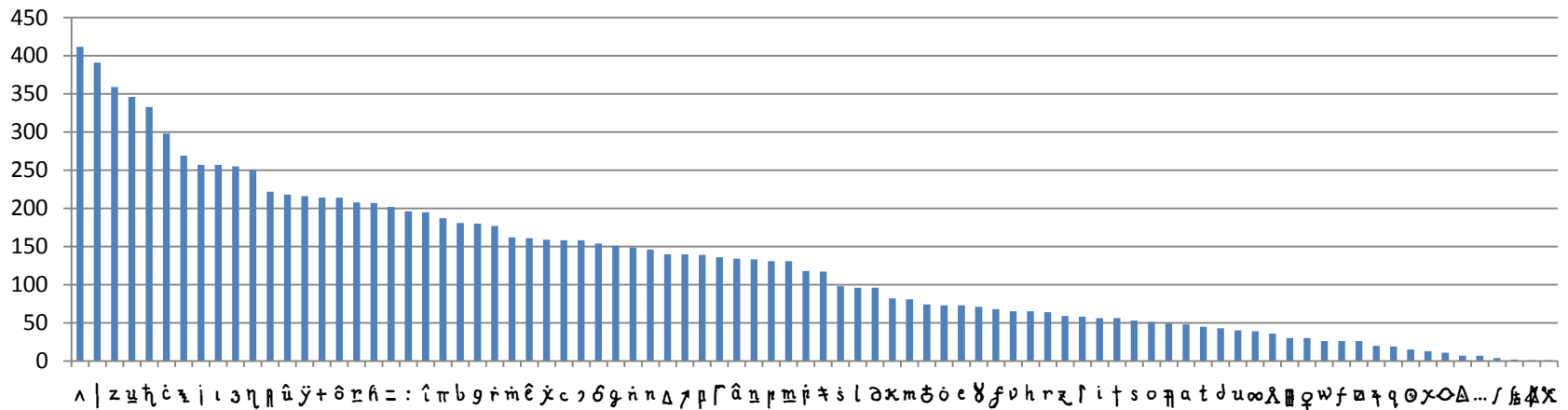
typed as:

pi oh j v hd tri arr
eh three c. ah ni arr
lam uh b lip uu r o..
zs

~90 cipher letters
(plus capitals)

Initially typed 16 pp
(10,840 letters).

Letter Frequencies



digraphs:

99

ċ : 66

h 49

: u 48

z R 44

trigraphs:

2 h 47

č : u 23

η ν η 22

ਯੂ ਚ ਫ਼ 18

h c | 17

tendencies:

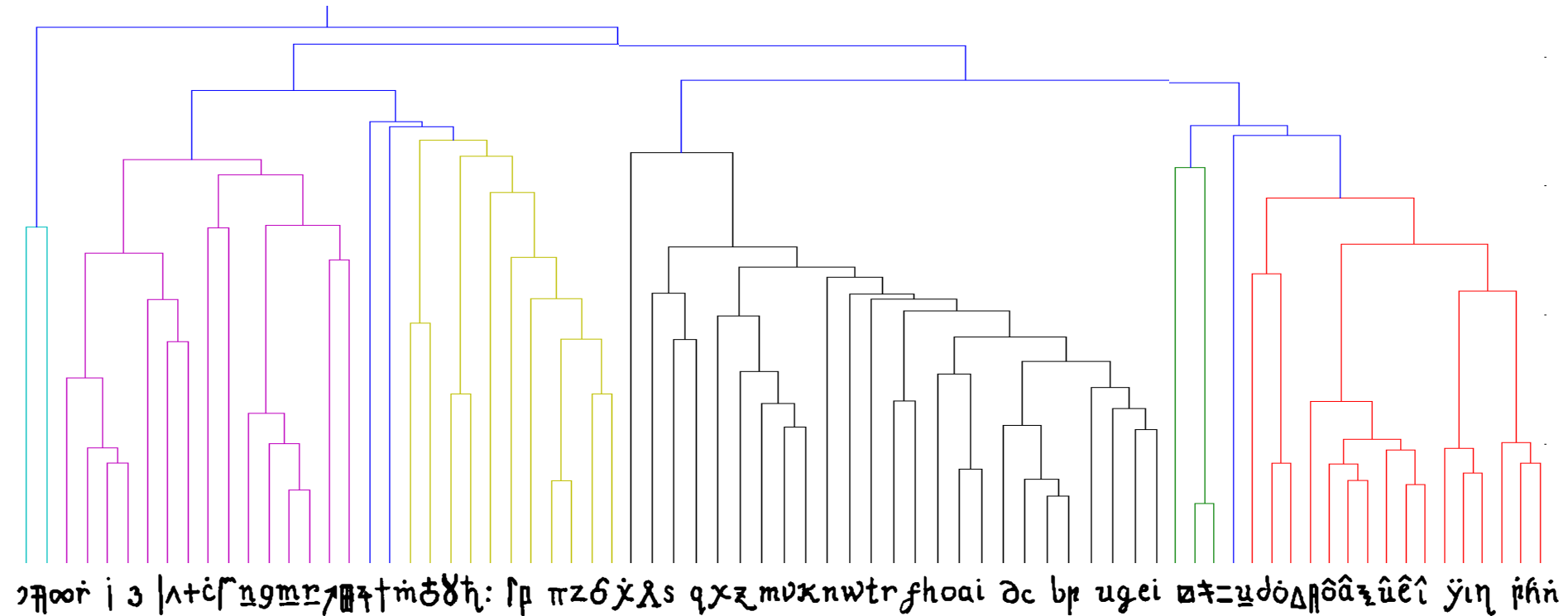
â, ê, î, ô, û followed by **3** and **j**

 $\hat{a}, \hat{e}, \hat{i}, \hat{o}, \hat{u}$ preceded by z and π

Clustering of Cipher Letters

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

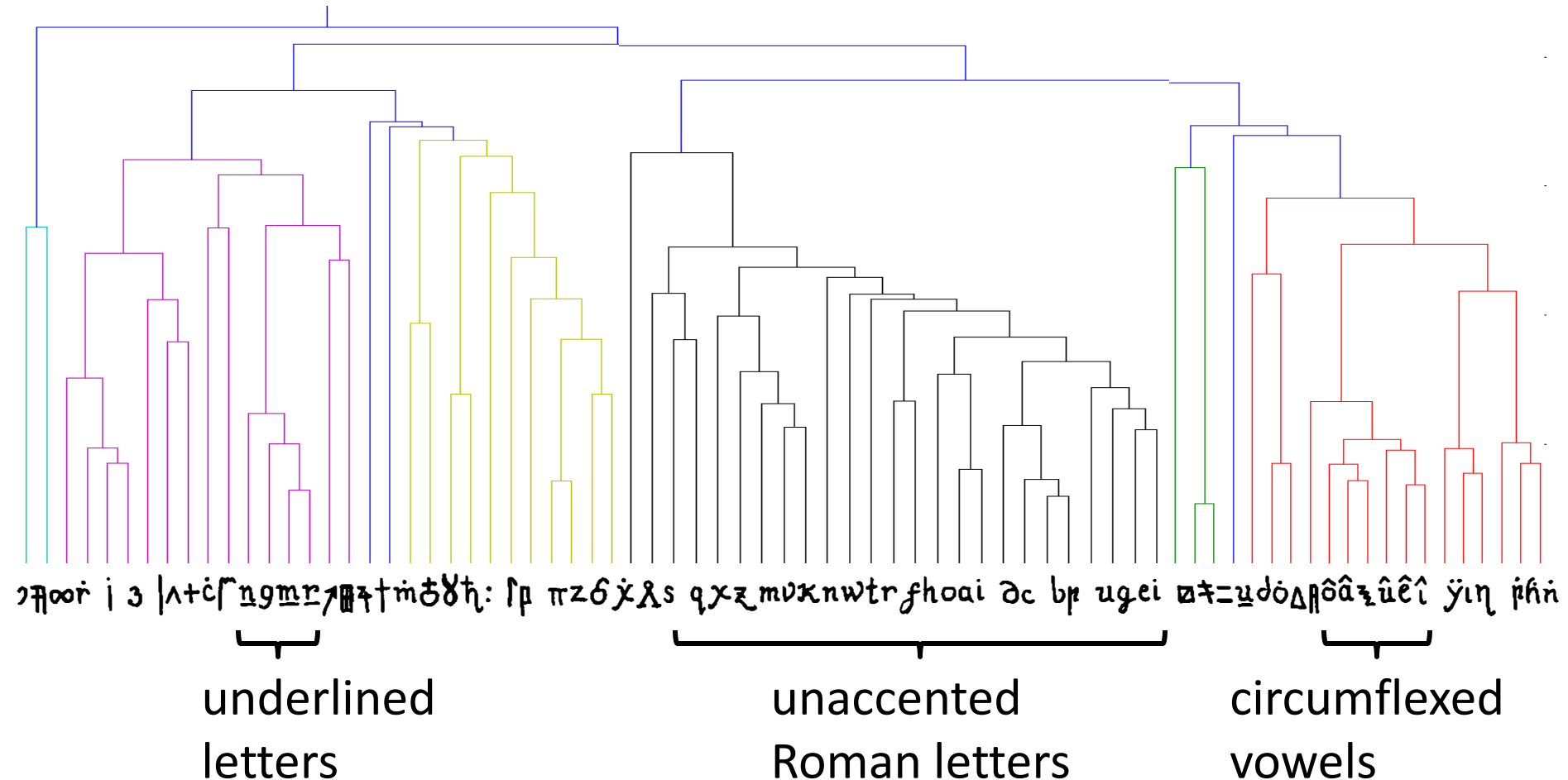
Scipy software



Clustering of Cipher Letters

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

Scipy software



thanks Jon Graehl

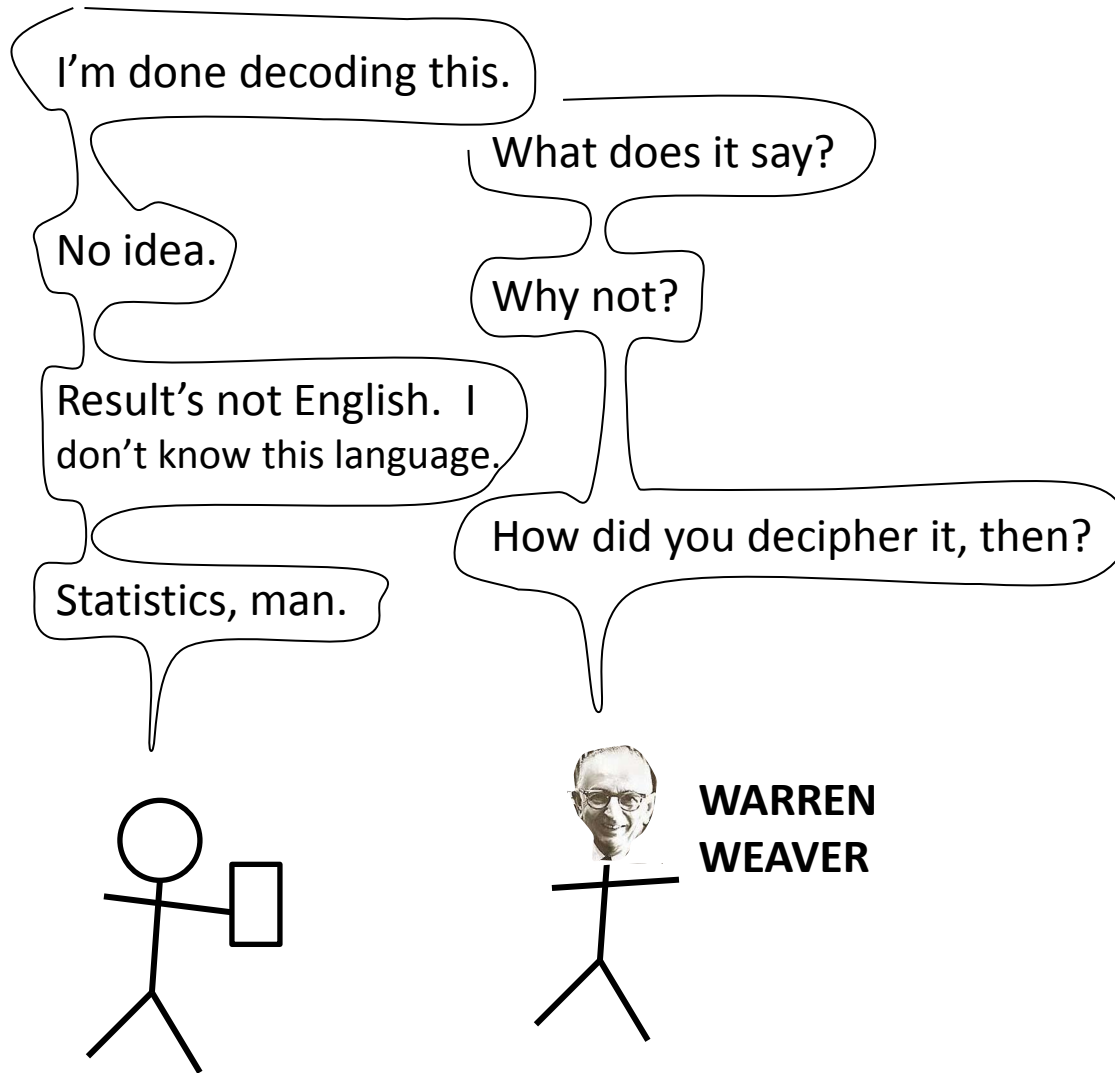
EM clustering (10 classes, a-j)

=====			c sqp	0.10	e bar	0.04	g tri	0.04	i o	0.02	
a c.	0.17		c nee	0.08	e ah	0.03	g grl	0.03	i r	0.02	
a bas	0.16		c o..	0.06	e grr	0.03	g h.	0.03	i s	0.02	
a ru	0.15		c :	0.05	e uh	0.03	g d	0.02	i t	0.02	
a mu	0.12		c cross	0.05	e hk	0.02	g ns	0.02	i u	0.02	
a g	0.10		c e	0.05	e ih	0.02	=====			i iot	0.01
a nu	0.06		c s.	0.05	=====			h j	0.22	i lam	0.01
a sqi	0.04		c i	0.04	f arr	0.17	h three	0.18	i q	0.01	
a bar	0.03		c lip	0.04	f lam	0.08	h r.	0.15	i w	0.01	
a lam	0.03		c tri	0.04	f ru	0.08	h hd	0.14	=====		
a j	0.02		c x.	0.04	f g	0.07	h bar	0.07	j z	0.11	
a oh	0.02		c ...	0.03	f del	0.06	h lam	0.07	j bar	0.10	
a plus	0.02		c tri..	0.03	f bar	0.05	h inf	0.04	j m.	0.09	
a r	0.02		c n	0.02	f mu	0.05	h plus	0.03	j c.	0.07	
a ah	0.01		c l	0.01	f nu	0.05	h y..	0.02	j plus	0.07	
a d	0.01		c s	0.01	f plus	0.05	h c.	0.01	j pi	0.06	
a zzz	0.01		=====			f grc	0.04	h cross	0.01	j hd	0.05
=====			d :	0.20	f sqi	0.04	h ns	0.01	j mal	0.04	
b iot	0.13		d z	0.17	f sqp	0.04	=====			j s.	0.04
b ni	0.13		d hd	0.12	f x.	0.04	i b	0.09	j sqp	0.04	
b ns	0.13		d pi	0.11	f oh	0.03	i c	0.08	j x.	0.04	
b y..	0.11		d bar	0.08	f three	0.03	i gs	0.08	j g	0.03	
b h.	0.09		d lam	0.07	f ah	0.02	i p	0.08	j del	0.02	
b n.	0.09		d del	0.05	f r.	0.02	i n	0.07	j gam	0.02	
b ki	0.08		d x.	0.05	f zzz	0.02	i ds	0.05	j h.	0.02	
b p.	0.06		d gam	0.03	f c.	0.01	i k	0.05	j lam	0.02	
b tri	0.06		d s.	0.02	f hk	0.01	i l	0.05	j sqi	0.02	
b o.	0.04		d sqi	0.01	=====			i f	0.04	j three	0.02
b ih	0.02		d sqp	0.01	g zzz	0.16	i a	0.03	j bas	0.01	
b fem	0.01		=====			g grr	0.14	i h	0.03	j j	0.01
b no	0.01		e uu	0.50	g uh	0.13	i i	0.03	j n.	0.01	
b uh	0.01		e lam	0.14	g ih	0.11	i m	0.03	j nu	0.01	
=====			e oh	0.06	g oh	0.10	i v	0.03	j p.	0.01	
c sqi	0.11		e eh	0.05	g eh	0.09	i zs	0.03	j ru	0.01	
c lam	0.10		e zzz	0.05	g ah	0.08	i e	0.02			

circumflexes

unaccented letters

Plaintext Language is Unknown



Knight et al [2006]:

Method 1:

“Universal” laws of human language

Method 2:

Brute-force decipher, versus all known languages

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

κ m û r: p z i ô f | y ʝ h ê j h z i λ n π â z b Δ g z =
i j l z u p q ç λ â r g κ λ h π r h π λ z i n p î | Δ r δ π λ a
= g z w π y ê c Δ r ô Δ + b z η r i x y j î r z u f λ z
π x j ʝ r = | û λ s x m δ m | z η g â | κ h = λ h | l x ô
ø | : r î λ b i f u m y j z v z â j x ʝ r m π i z h λ c ô o g g
z û + r x π n x ê z g h λ h π i h π λ î z t n x | r ô y m â
+ h h r z ô z n b s η + : z r κ r p δ π h Δ c û λ g = n z κ
p z ∞ n z π z f h n r π z ê y n g = r π g ô Δ z n z π | κ
π η j i x y r î g π u λ b g λ i t b ô û | π h z ô λ e z η ô
| û r c | z q δ λ b η h m θ

κ f n g l κ n a c b f z m κ
l b u v c g h t r h b κ g n κ n
f g g n κ b g b e c b ...

Decipher against
80 plaintext languages.

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

κ m û r: p z i ô f | y ʝ h ê j h z i λ n π â z b Δ g z =
i j l z u p q ç λ â r g κ λ h π r h π λ z i n p î | Δ r δ π λ a
= g z w π y ê c Δ r δ Δ + b z η r i x y j î r z u f λ z
π x j ʝ r = f | û λ s x m δ m | z η g â | κ h = λ h | l x ô
ø f: r î λ b i f u m y j z v z â j x ʝ r m π i z h λ c δ ô g g
z û + r x π n x ê z g h λ h π i h π λ î z t n x | r ô y m â
+ h h r z ô z n b s η +: z r κ r p δ π h Δ ç û λ g = n z κ
p z ∞ n z π z f h n r π z ê y n g = r π g δ Δ z n z π | κ
π η j i x y r î g π u λ b g λ i t b δ û f π h z ô λ e z η ô
| û r c f z q δ λ b η h m θ

κ f n g l κ n a c b f z m κ
l b u v c g h t r h b κ g n κ n
f g g n κ b g b e c b ...

D
80
FAIL
rst
languages.

Second Decipherment Approach

Homophonic cipher,
e.g.:

A = ʒ i l y r

B = û

C = ô ñ

D = ʈ

E = ʃ ɸ Δ * f î ʒ 3

F = p

G = ȳ

etc.



κ m û r : p z i ô ɸ | ȳ ʌ ħ ê j ħ ʒ i λ n π â ʒ b Δ g z =
i ʒ l z u p ɸ ç λ â r g κ λ ħ ʈ r ħ ʈ λ ʒ i n p î | Δ r ô ʈ λ λ a
= g z w π ȳ ê c Δ r ô Δ + b z η r i ʃ ȳ j î r z u ɸ λ ʒ
π * j ʈ r = ʈ | û λ s * m δ m | ʒ η g â | κ ħ = λ ħ | l ʃ ô
ø ʈ : r î λ b i ʈ u m ȳ j z v z â j x ʌ r m π i z ħ λ c ô ô g g
z û + p * ʈ ñ ʃ ê ʒ g ħ λ ħ ʈ i ħ ʈ λ î ʒ t n * | r ô ȳ m â
+ ħ ħ r z ô ʒ ñ b s η + : ʒ r κ r p ô ʈ ħ Δ ç û λ g = ñ z κ
p ʒ ∞ n z ʈ ʒ ɸ ħ ñ r π ʒ ê ȳ ñ g = r π g ô Δ ʒ n z ʈ | κ
π η j i ʃ ȳ r î g π u λ b g λ i ʈ b ô û ʈ ʈ ħ ʒ ô λ e z η ô
| û r c ʈ ʒ ɸ ô λ b η ħ m ð

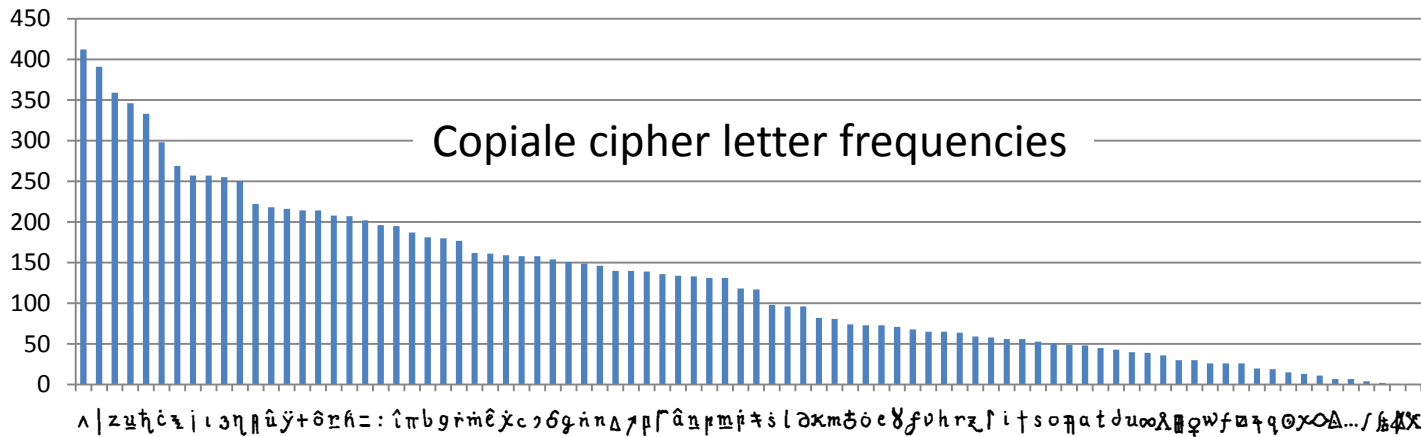
Homophonic Cipher

Assign cipher-letter types based on plaintext-letter frequencies

→ uniform distribution of cipher types!

→ harder to crack

but recall:



**on the
other hand:**



Homophonic Cipher

- Let's try computer attack, assuming plaintext English, then Latin, then French ...
 - [Knight et al 06] method, using letter n-gram LM
- But suppose we fail:
 - was it because the cipher is not homophonic?
 - or because homophonic ciphers at this length cannot be reliably solved by our method?
- Reality check:
 - first set up synthetic, actual homophonic cipher
 - yes, we can solve it! (and ID the plaintext language)

Homophonic Cipher

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

FAIL

Extremely slight numerical preference
for German ... random noise?

What Next?

- Concentrate on German
 - book located in Germany
 - slight computer preference for German
 - inscription “Philipp 1866”
- Roll up sleeves

Cipher Characteristics

digraphs:

ʁ ʰ 99

č : 66

ʰ ^ 49

: ȳ 48

z ʀ 44

trigraphs:

ʁ ʰ ^ 47

č : ȳ 23

ʱ ʁ ʰ 22

ȳ ʁ ʰ 18

ʰ č | 17

tendencies:

â, ê, î, ô, û followed by ʒ and j

â, ê, î, ô, û preceded by z and π



?



?

**should appear
adjacent in German text**

Make full digraph table for cipher and for German

Digraph Tables

Key Observation #1

In Copiale, ʔ almost always followed by ᵀ

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

So guess: ʔ = C, ᵀ = H

One Thing Leads to Another

$\text{ʁ}^{\text{h}} = \text{CH} \rightarrow \text{ʁ}^{\text{h}}\text{ʌ} = \text{CHT} \rightarrow \text{ʌ} = \text{T} ?$

Each step is guesswork.

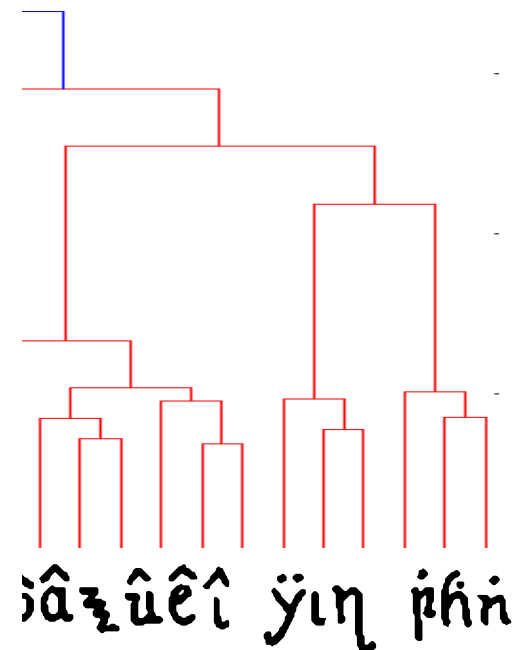
Must be willing to retract.

Weird task, not knowing German.

No longer care what the book says.

Cluster diagram crucial:

$\ddot{y} = \text{I} \rightarrow \text{ɪ} = \text{I} , \eta = \text{I}$



Spring Break 2011

c aeiou fpy dlmrztbvw hkngrs j

german ^{aus} ub

8 f
 1 h
 2 z
 3 o
 4 u
 5 i
 6 e
 7 a
 8 o
 9 u
 10 i
 11 e
 12 a
 13 o
 14 u
 15 i
 16 e
 17 a
 18 o
 19 u
 20 i
 21 e
 22 a
 23 o
 24 u
 25 i
 26 e
 27 a
 28 o
 29 u
 30 i
 31 e
 32 a
 33 o
 34 u
 35 i
 36 e
 37 a
 38 o
 39 u
 40 i
 41 e
 42 a
 43 o
 44 u
 45 i
 46 e
 47 a
 48 o
 49 u
 50 i
 51 e
 52 a
 53 o
 54 u
 55 i
 56 e
 57 a
 58 o
 59 u
 60 i
 61 e
 62 a
 63 o
 64 u
 65 i
 66 e
 67 a
 68 o
 69 u
 70 i
 71 e
 72 a
 73 o
 74 u
 75 i
 76 e
 77 a
 78 o
 79 u
 80 i
 81 e
 82 a
 83 o
 84 u
 85 i
 86 e
 87 a
 88 o
 89 u
 90 i
 91 e
 92 a
 93 o
 94 u
 95 i
 96 e
 97 a
 98 o
 99 u
 100 i
 101 e
 102 a
 103 o
 104 u
 105 i
 106 e
 107 a
 108 o
 109 u
 110 i
 111 e
 112 a
 113 o
 114 u
 115 i
 116 e
 117 a
 118 o
 119 u
 120 i
 121 e
 122 a
 123 o
 124 u
 125 i
 126 e
 127 a
 128 o
 129 u
 130 i
 131 e
 132 a
 133 o
 134 u
 135 i
 136 e
 137 a
 138 o
 139 u
 140 i
 141 e
 142 a
 143 o
 144 u
 145 i
 146 e
 147 a
 148 o
 149 u
 150 i
 151 e
 152 a
 153 o
 154 u
 155 i
 156 e
 157 a
 158 o
 159 u
 160 i
 161 e
 162 a
 163 o
 164 u
 165 i
 166 e
 167 a
 168 o
 169 u
 170 i
 171 e
 172 a
 173 o
 174 u
 175 i
 176 e
 177 a
 178 o
 179 u
 180 i
 181 e
 182 a
 183 o
 184 u
 185 i
 186 e
 187 a
 188 o
 189 u
 190 i
 191 e
 192 a
 193 o
 194 u
 195 i
 196 e
 197 a
 198 o
 199 u
 200 i
 201 e
 202 a
 203 o
 204 u
 205 i
 206 e
 207 a
 208 o
 209 u
 210 i
 211 e
 212 a
 213 o
 214 u
 215 i
 216 e
 217 a
 218 o
 219 u
 220 i
 221 e
 222 a
 223 o
 224 u
 225 i
 226 e
 227 a
 228 o
 229 u
 230 i
 231 e
 232 a
 233 o
 234 u
 235 i
 236 e
 237 a
 238 o
 239 u
 240 i
 241 e
 242 a
 243 o
 244 u
 245 i
 246 e
 247 a
 248 o
 249 u
 250 i
 251 e
 252 a
 253 o
 254 u
 255 i
 256 e
 257 a
 258 o
 259 u
 260 i
 261 e
 262 a
 263 o
 264 u
 265 i
 266 e
 267 a
 268 o
 269 u
 270 i
 271 e
 272 a
 273 o
 274 u
 275 i
 276 e
 277 a
 278 o
 279 u
 280 i
 281 e
 282 a
 283 o
 284 u
 285 i
 286 e
 287 a
 288 o
 289 u
 290 i
 291 e
 292 a
 293 o
 294 u
 295 i
 296 e
 297 a
 298 o
 299 u
 300 i
 301 e
 302 a
 303 o
 304 u
 305 i
 306 e
 307 a
 308 o
 309 u
 310 i
 311 e
 312 a
 313 o
 314 u
 315 i
 316 e
 317 a
 318 o
 319 u
 320 i
 321 e
 322 a
 323 o
 324 u
 325 i
 326 e
 327 a
 328 o
 329 u
 330 i
 331 e
 332 a
 333 o
 334 u
 335 i
 336 e
 337 a
 338 o
 339 u
 340 i
 341 e
 342 a
 343 o
 344 u
 345 i
 346 e
 347 a
 348 o
 349 u
 350 i
 351 e
 352 a
 353 o
 354 u
 355 i
 356 e
 357 a
 358 o
 359 u
 360 i
 361 e
 362 a
 363 o
 364 u
 365 i
 366 e
 367 a
 368 o
 369 u
 370 i
 371 e
 372 a
 373 o
 374 u
 375 i
 376 e
 377 a
 378 o
 379 u
 380 i
 381 e
 382 a
 383 o
 384 u
 385 i
 386 e
 387 a
 388 o
 389 u
 390 i
 391 e
 392 a
 393 o
 394 u
 395 i
 396 e
 397 a
 398 o
 399 u
 400 i
 401 e
 402 a
 403 o
 404 u
 405 i
 406 e
 407 a
 408 o
 409 u
 410 i
 411 e
 412 a
 413 o
 414 u
 415 i
 416 e
 417 a
 418 o
 419 u
 420 i
 421 e
 422 a
 423 o
 424 u
 425 i
 426 e
 427 a
 428 o
 429 u
 430 i
 431 e
 432 a
 433 o
 434 u
 435 i
 436 e
 437 a
 438 o
 439 u
 440 i
 441 e
 442 a
 443 o
 444 u
 445 i
 446 e
 447 a
 448 o
 449 u
 450 i
 451 e
 452 a
 453 o
 454 u
 455 i
 456 e
 457 a
 458 o
 459 u
 460 i
 461 e
 462 a
 463 o
 464 u
 465 i
 466 e
 467 a
 468 o
 469 u
 470 i
 471 e
 472 a
 473 o
 474 u
 475 i
 476 e
 477 a
 478 o
 479 u
 480 i
 481 e
 482 a
 483 o
 484 u
 485 i
 486 e
 487 a
 488 o
 489 u
 490 i
 491 e
 492 a
 493 o
 494 u
 495 i
 496 e
 497 a
 498 o
 499 u
 500 i
 501 e
 502 a
 503 o
 504 u
 505 i
 506 e
 507 a
 508 o
 509 u
 510 i
 511 e
 512 a
 513 o
 514 u
 515 i
 516 e
 517 a
 518 o
 519 u
 520 i
 521 e
 522 a
 523 o
 524 u
 525 i
 526 e
 527 a
 528 o
 529 u
 530 i
 531 e
 532 a
 533 o
 534 u
 535 i
 536 e
 537 a
 538 o
 539 u
 540 i
 541 e
 542 a
 543 o
 544 u
 545 i
 546 e
 547 a
 548 o
 549 u
 550 i
 551 e
 552 a
 553 o
 554 u
 555 i
 556 e
 557 a
 558 o
 559 u
 560 i
 561 e
 562 a
 563 o
 564 u
 565 i
 566 e
 567 a
 568 o
 569 u
 570 i
 571 e
 572 a
 573 o
 574 u
 575 i
 576 e
 577 a
 578 o
 579 u
 580 i
 581 e
 582 a
 583 o
 584 u
 585 i
 586 e
 587 a
 588 o
 589 u
 590 i
 591 e
 592 a
 593 o
 594 u
 595 i
 596 e
 597 a
 598 o
 599 u
 600 i
 601 e
 602 a
 603 o
 604 u
 605 i
 606 e
 607 a
 608 o
 609 u
 610 i
 611 e
 612 a
 613 o
 614 u
 615 i
 616 e
 617 a
 618 o
 619 u
 620 i
 621 e
 622 a
 623 o
 624 u
 625 i
 626 e
 627 a
 628 o
 629 u
 630 i
 631 e
 632 a
 633 o
 634 u
 635 i
 636 e
 637 a
 638 o
 639 u
 640 i
 641 e
 642 a
 643 o
 644 u
 645 i
 646 e
 647 a
 648 o
 649 u
 650 i
 651 e
 652 a
 653 o
 654 u
 655 i
 656 e
 657 a
 658 o
 659 u
 660 i
 661 e
 662 a
 663 o
 664 u
 665 i
 666 e
 667 a
 668 o
 669 u
 670 i
 671 e
 672 a
 673 o
 674 u
 675 i
 676 e
 677 a
 678 o
 679 u
 680 i
 681 e
 682 a
 683 o
 684 u
 685 i
 686 e
 687 a
 688 o
 689 u
 690 i
 691 e
 692 a
 693 o
 694 u
 695 i
 696 e
 697 a
 698 o
 699 u
 700 i
 701 e
 702 a
 703 o
 704 u
 705 i
 706 e
 707 a
 708 o
 709 u
 710 i
 711 e
 712 a
 713 o
 714 u
 715 i
 716 e
 717 a
 718 o
 719 u
 720 i
 721 e
 722 a
 723 o
 724 u
 725 i
 726 e
 727 a
 728 o
 729 u
 730 i
 731 e
 732 a
 733 o
 734 u
 735 i
 736 e
 737 a
 738 o
 739 u
 740 i
 741 e
 742 a
 743 o
 744 u
 745 i
 746 e
 747 a
 748 o
 749 u
 750 i
 751 e
 752 a
 753 o
 754 u
 755 i
 756 e
 757 a
 758 o
 759 u
 760 i
 761 e
 762 a
 763 o
 764 u
 765 i
 766 e
 767 a
 768 o
 769 u
 770 i
 771 e
 772 a
 773 o
 774 u
 775 i
 776 e
 777 a
 778 o
 779 u
 780 i
 781 e
 782 a
 783 o
 784 u
 785 i
 786 e
 787 a
 788 o
 789 u
 790 i
 791 e
 792 a
 793 o
 794 u
 795 i
 796 e
 797 a
 798 o
 799 u
 800 i
 801 e
 802 a
 803 o
 804 u
 805 i
 806 e
 807 a
 808 o
 809 u
 810 i
 811 e
 812 a
 813 o
 814 u
 815 i
 816 e
 817 a
 818 o
 819 u
 820 i
 821 e
 822 a
 823 o
 824 u
 825 i
 826 e
 827 a
 828 o
 829 u
 830 i
 831 e
 832 a
 833 o
 834 u
 835 i
 836 e
 837 a
 838 o
 839 u
 840 i
 841 e
 842 a
 843 o
 844 u
 845 i
 846 e
 847 a
 848 o
 849 u
 850 i
 851 e
 852 a
 853 o
 854 u
 855 i
 856 e
 857 a
 858 o
 859 u
 860 i
 861 e
 862 a
 863 o
 864 u
 865 i
 866 e
 867 a
 868 o
 869 u
 870 i
 871 e
 872 a
 873 o
 874 u
 875 i
 876 e
 877 a
 878 o
 879 u
 880 i
 881 e
 882 a
 883 o
 884 u
 885 i
 886 e
 887 a
 888 o
 889 u
 890 i
 891 e
 892 a
 893 o
 894 u
 895 i
 896 e
 897 a
 898 o
 899 u
 900 i
 901 e
 902 a
 903 o
 904 u
 905 i
 906 e
 907 a
 908 o
 909 u
 910 i
 911 e
 912 a
 913 o
 914 u
 915 i
 916 e
 917 a
 918 o
 919 u
 920 i
 921 e
 922 a
 923 o
 924 u
 925 i
 926 e
 927 a
 928 o
 929 u
 930 i
 931 e
 932 a
 933 o
 934 u
 935 i
 936 e
 937 a
 938 o
 939 u
 940 i
 941 e
 942 a
 943 o
 944 u
 945 i
 946 e
 947 a
 948 o
 949 u
 950 i
 951 e
 952 a
 953 o
 954 u
 955 i
 956 e
 957 a
 958 o
 959 u
 960 i
 961 e
 962 a
 963 o
 964 u
 965 i
 966 e
 967 a
 968 o
 969 u
 970 i
 971 e
 972 a
 973 o
 974 u
 975 i
 976 e
 977 a
 978 o
 979 u
 980 i
 981 e
 982 a
 983 o
 984 u
 985 i
 986 e
 987 a
 988 o
 989 u
 990 i
 991 e
 992 a
 993 o
 994 u
 995 i
 996 e
 997 a
 998 o
 999 u
 1000 i

1 P
 2 P
 3 P
 4 P
 5 P
 6 P
 7 P
 8 P
 9 P
 10 P
 11 P
 12 P
 13 P
 14 P
 15 P
 16 P
 17 P
 18 P
 19 P
 20 P
 21 P
 22 P
 23 P
 24 P
 25 P
 26 P
 27 P
 28 P
 29 P
 30 P
 31 P
 32 P
 33 P
 34 P
 35 P
 36 P
 37 P
 38 P
 39 P
 40 P
 41 P
 42 P
 43 P
 44 P
 45 P
 46 P
 47 P
 48 P
 49 P
 50 P
 51 P
 52 P
 53 P
 54 P
 55 P
 56 P
 57 P
 58 P
 59 P
 60 P
 61 P
 62 P
 63 P
 64 P
 65 P
 66 P
 67 P
 68 P
 69 P
 70 P
 71 P
 72 P
 73 P
 74 P
 75 P
 76 P
 77 P
 78 P
 79 P
 80 P
 81 P
 82 P
 83 P
 84 P
 85 P
 86 P
 87 P
 88 P
 89 P
 90 P
 91 P
 92 P
 93 P
 94 P
 95 P
 96 P
 97 P
 98 P
 99 P
 100 P

vowels: u ɔ ɛ ɪ ʏ
 Δ o p i h
 3 m c r + = #

need: f g y l m z
 (rare on
 german)
 u f sp
 a f pr
 r f pe
 f

1 Δ
 a o i
 v w
 k g j
 1 Δ
 a o i
 v w
 k g j
 1 Δ
 a o i
 v w
 k g j

1 Δ
 a o i
 v w
 k g j
 1 Δ
 a o i
 v w
 k g j
 1 Δ
 a o i
 v w
 k g j

* der ✓ eht 2 h A
 und ✓ e i u
 ein ✓ s l h i t
 ung ✓ ich 2 h
 eht ✓ ich 2 h
 ich ✓ che 2 h
 sch ✓ t h
 che ✓ t h
 ech ✓ t h
 die ✓ s a h
 rec ds 2 h
 ine eil t h
 gen t c l n
 eit ✓ t e l i u
 ver ti l i
 hen der 2 h
 lic der 2 h
 ten ew u a i
 rei sch h 2 h
 nte st h k l
 auf ich l 2 h
 ede ein h y t
 and ✓ t 2 h
 den die 2 h
 run die 2 h
 tar die 2 h
 frei e 2 h
 sei e 2 h
 hte h t m h
 hei und = n t
 nre z m a i c
 ens i h l c
 men tot 2 h
 hat o h m z
 ere i c
 das a k t m
 rde ste h a u
 nte h c
 nge sch h 2 h
 tie c i e
 ore che 2 h
 ede t l a c

Spring Break 2011

German letters

Cipher letters, in groups

Grid

vowels: u ʔ ɛ ɪ ɔ ɯ

$$\Delta \hat{O} \hat{P} \hat{n} \hat{h}$$
$$\gamma \frac{1}{m} \dot{p} = 1 \neq 0$$

need: f g y l m z

-(rapidly / slowly)

uf sp
a.f pr
r.f pr
ff

aoi
vw k g

Attard	Ke	Ge
Sydney	Ku	Gr
Albany		St
		Gr
		na

$$Z_U \quad \{v_i\} \in \{nrd\}$$

a u f o

$$\begin{array}{cc} u & v \\ u & v \end{array}$$

5. Find $\frac{dy}{dx}$ if $y = \sin^{-1} x$

318

31.

1. γ ω

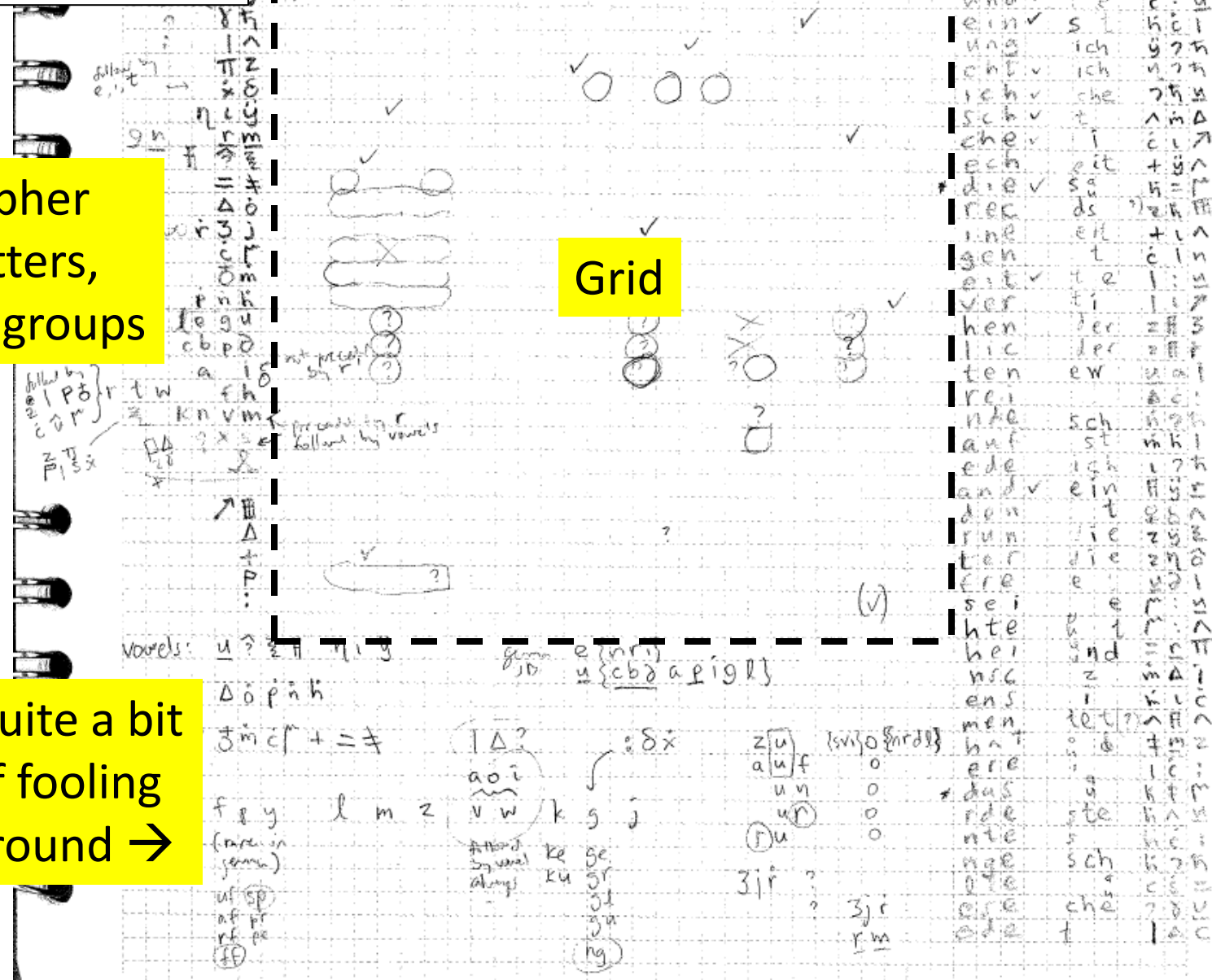
der	✓	cht	2H	Λ
und	✓	e	c:	u
ein	✓	s	h	i
una		ich	g	7H
cht	✓	ich	n	7H
ich	✓	che	7H	u
sch	✓	t	Λ	MA
che	✓	i	c	L
ech		ait	+g	Λ
die	✓	sä	+h	Λ
rec		ds	7H	Λ
ne		et	+i	Λ
gen		t	c	Λ
oit	✓	t	i	Λ
ver		ti	i	7
hen		ter	=H	3
lic		der	2H	Λ
ten		ew	u	Λ
re			Λ	c
na		sch	h	7H
anf		st	n	h
ede		igh	i	7H
and	✓	ein	H	5
den		t	2H	Λ
run		die	z	g
tar		die	z	g
ere		e	u	Λ
sei		e	m	Λ
hte		u	m	Λ
hei		und	=	Λ
nsc		z	m	Λ
ens		i	i	Λ
men		te	Λ	Λ
hat		o	Λ	Λ
erie		g	Λ	Λ
das		ste	h	Λ
nde			h	Λ
nte		sch	h	Λ
ge			h	Λ
le		che	7H	Λ
ore			7H	Λ
ore		t	Λ	Λ

Spring Break 2011

German letters

Cipher letters, in groups

Quite a bit
of fooling
around →



Problems Remain

Good progress:

- But very few SCH coming out
- No mappings yet for unaccented Roman letters

Decipherment with ? for cipher letters still without mappings:

?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

κ m û r: p z i ô f | y ʔ h ê j h z i λ n π â z b Δ g z =
i j l z u p q ç λ â r g κ λ h p r h p λ z i n p î | Δ r δ p λ a
= g z w π y ê c Δ r ð Δ + b z η r i x y j î r z u f λ z
π x j ʔ ð r = f | û λ s x m δ m | z η g â | κ h = λ h | l x ô
ø f: r î λ b i f u m y j z v z â j x ʔ r m π i z h λ c ð ô g g
z û + p x p n x ê z g h λ h p i h p λ î z t n x | r ô y m â
+ h h r z ô z n b s η +: z r κ r p δ p h Δ c û λ g = n z κ
p z ∞ n z p z f h n r π z ê y n g = r π g ð Δ z n z p | κ
π η j i x y r î g π u λ b g λ i t b δ û f p h z ô λ e z η ô
| û r c f z q δ λ b η h m ð

Actually, those are space bars

Use Online German-English Translator www.freetranslation.com



êîŕê³ îþ†gη∧: → Erder Abschnittl → Erder Abschnittl

Use Online German-English Translator www.freetranslation.com

ûĩŕê³ ïþ†gη∧: → Erder Abschnittl → Erder Abschnittl
Erder Abschnita → Erder Abschnita
Erder Abschnittb → Erder Abschnittb
Erder Abschnittc → Erder Abschnittc
...
Erder Abschnitt → Erder Section

Cipher “:”
maps to
German L

Cipher “:”
maps to
German T

Another Strange Letter: †

†AFLNER, †NUPFTUCHS, and GESELL†AFLT

Tried substituting A-Z for †

Found GESELLSHAFLT in dictionary, †=**SCH**

Led to others multi-letter substitutions:

û†ê³ ï†gη^: → ~~Er~~der Abschnitt → First Section
st

†=**ST**

German Native Edits Decipherment

eÿʀt+nícĥzēgṽpôṛfʀíç|nñŋčđʦĥzčÿʃônrz mēxʔgôṽ

hypothesis: is mache ebenfalls wilhuhrlise bewegunge

correction: ich mache ebenfalls wilkührliche bewegungenn

απΔʃb+ιλgʒʃjvĥgzēzgsĥñmπv

hypothesis: dos mit der andern hand

correction: doch mit der andern hand

rz|ŋλ:îg|eṽÿñîjz mŋ+:λbλôpñcđʀ+ιλʃπṽgʒĥ=+ṽc

hypothesis: dritlens einer n mlt tobach mit de daume

correction: drittens einer ??? tobach mit demm daumenen

ʔmπlzṽt+ιλ:ôç|ṽcʃÿgʔzēszʃʀĥçιrʔδṽðĥññzb

hypothesis: und de mitlelde finger der linche hand

correction: und demm mittelsten finger der linchenn hand

Substitution Table

German	Ciphertext
A	þ ñ ħ ǵ
Ä	ǵ
B	þ
C	ʝ
D	π z
E	â ê î ô û ʀ ʒ
F	ƿ
G	δ ȳ
H	ħ ʝ
I	ÿ ŋ ı
J	ʒ
K	ʝ
L	ċ
M	+
N	ṁ ṛ ṅ ɡ
O	Δ ò
Ö	⊠
P	ɔ

German	Ciphertext
R	ř ʒ ı
S	⊠
T	^
U	= ʒ
Ü	⌘
V	⦿
W	ṁ
X	ƒ
Y	∞
Z	š
SCH	†
SS	⊠
ST	ſ
CH	↗
repeat	:
EN / EM	u
space	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

German	Ciphertext
A	þ ñ ĥ ♀
Ä	♀
B	þ
C	ʝ
D	π z
E	â ê î ô û þ ʞ
F	ƒ
G	δ ȳ
H	ħ ʝ
I	ÿ η ι
J	ȳ
K	ʝ
L	ċ
M	+
N	ṁ ṛ ṅ 9
O	Δ ô
Ö	⊞
P	ɖ

German	Ciphertext
R	ř ʒ
S	⊞
T	^
U	= ʞ
Ü	⌘
V	⊞
W	ṁ
X	ƒ
Y	∞
Z	š
SCH	†
SS	⊞
ST	ſ
CH	↗
repeat	:
EN / EM	u
space	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

Substitution Table

Cipher letters in **RED** decode ambiguously ...

But only with respect to modern German, not old German.

No entries for large characters:
 Ɑ, Ɱ, Δ, ʞ, ⊞, ⌘, ƒ, ɖ, Π

Copiale Decipherment

lit:mz a/bl

$$v \dot{x} \approx |\hat{v} \wedge \dot{x}| \approx \dot{x} / \omega_n$$
$$\pi \circ \nu \upharpoonright_{\Delta} \upharpoonright_{\hat{e} \cup \hat{a}} = \upharpoonright_{\hat{a} \cup \hat{b}} \circ \underline{u} \circ \tau$$
$$\delta A \hbar \Xi_1 + \delta j \eta \wedge \hbar \hat{\eta} \dot{c} f.$$

cûîfê3tîp†9η∧:κ

$\partial \hat{x} \hat{u} \hat{t} \hat{e} \eta + \pi \hat{r} \hat{\pi} \hat{m} \wedge \hat{i} \hat{s} : \ddot{y} \hat{\gamma} \wedge \hat{l} \hat{o} \hat{o} \hat{q} z i \hat{i} \hat{x} \hat{a} | \hat{e} \hat{c} : \underline{u} \partial.$

$$f \hat{u} r \hat{o} j \kappa \wedge \iota \wedge = \dot{c} z.$$
$$m\omega\hat{r}\hat{a}+\Delta g\ddot{u}xz\hat{o}_3m\hat{n}\hat{x}\hat{\Gamma}\hat{n}\hat{h}\hat{t}+\hat{i}f.$$

$\kappa\pi\bar{u}r:\mu z i \hat{o} f | \ddot{y} : h \hat{e} | h_{\Sigma} \wedge n \pi \hat{a} z b \Delta g z = j / l z u r q \dot{\epsilon} \lambda \hat{a} r g \kappa \wedge h$

$\pi r^h \pi \lambda z i n p i \Delta r \delta \rho \lambda a = g z w \pi \ddot{y} \hat{e} c \Delta r \delta \Delta + b z \eta r i \dot{x} \ddot{y} i \hat{r} z u f \lambda z$

πκί|∂ř=Γ|ûΛ\$κ̄mδm|εηgâ|κh=Λh|[đøøΓ:r̂Λbifumÿ|zv

$\hat{z} \hat{\alpha}_j x_j r_m \pi i z h l c \delta o g g z \hat{u} + r_+ p n \dot{x} \hat{e} z g h \wedge \tilde{h} p i \tilde{h} p \wedge \hat{i} s t n \ddagger | r o y$

$\underline{m}\bar{a}+h\bar{h}\underline{r}z\bar{o}\bar{z}\bar{n}\bar{b}\bar{s}\bar{n}+:\bar{x}\bar{r}\bar{x}\bar{r}\bar{p}\bar{o}\bar{h}\bar{\Delta}\bar{c}\bar{u}\bar{g}=\bar{n}\bar{z}\bar{x}\bar{p}\bar{x}\bar{o}\bar{o}\bar{n}\bar{z}\bar{a}\bar{z}\bar{f}\bar{h}\bar{n}\bar{r}\bar{\pi}$

zēyng=rπgδΔznz||κπη|χÿrîgπuΛbgλ†bδûΓ#hδλc

$\alpha \in \Gamma_0(N)$

nîrîcî 7 eôphôrî 3 pî 6 û hî 3 â κ ◊ g s = 3 m ê j z u l i

g s m u o o l o n | o p e g u d o t r i d | z g t r x u g p i u x o n | r l b = n

Λέριμô3f:uα=rzLΔηrηθmηâzô||dîrflô8ôÿλvziêgçûη



 $\hat{r} \hat{s} \hat{e} \eta \wedge$
$$c_j = |\hat{a}^\dagger|_{\underline{b}m_\Delta} c_{j+2}.$$
$$hzi\eta\wedge\hat{o}g|ez\eta\hat{a}z\hat{A}n\pi\hat{a}z\hat{O}rs\ddot{r}\hat{o}\hat{i}z\ddot{t}m\hat{a}\ddot{y}\delta\ddot{u}l\ddot{i}=\ddot{n}\pi$$

z p̄s= n h̄ x̄ f r z p̄ | n ḡ x̄ | p̄ i g π c̄ ȳ o t̄ h̄ | ḡ b | i x̄ n g t̄ x̄ ȳ t̄ | ḡ c s̄ = b + n x̄ u ḡ

$$\dot{x} \in \mathbb{R}^n, \dot{u} \in \mathbb{R}^m.$$

Ἰπὸ τῆς ἡγεσίας τοῦ ἀρχιεπισκόπου Ἀθανασίου.

gesetz buchs

der hocheleuchte  e 

geheimer theil.

erster abschnitt

geheimer unterricht vor die gesellen.

erster titul.

ceremonien der aufnahme.

wenn die sicherheit der Δ durch den ältern




thürhüter besorget und die Δ vom dirigierenden λ

durch aufsetzung seines huths geöffnet ist wird der

candidat von dem jüngern thürhüter aus einem andern

zimmer abgeholt und bey der hand ein und vor des

dirigierenden  tisch geführt dieser fragt ihn:

erstlich ob er begehre zu werden

zweytens denen verordnungen der \odot sich

unterwerffen und ohne widerspenstigkeit die lehrzeit
ausstehen wolle.

drittens die ~~4~~ der ② gu verschweigen und dazu

auf das verbindlichste sich anheischig zu machen

gesinnet sey.

der kandidat antwortet ja.

Copiale Decipherment

lit:mz||bl
 0x3/11s3k37/wn
 π0j0hΔ7ē3câ=7λûb ① ur ② z

δh3ε+0jnpλhîηcf.

cûjē3tîp†gηλ:κ

0xûhēη+3r3mλî3:ÿ7λδ0jqziîxâ|ēc:uð.

fûrδjκλλ=cz.

m7hîâ+Δgÿuxz03mî3f7hî+îf.

κmûr:7zi0f|ÿ7hēj3λλnπâ3bΔgz=7λzup0cλâ7gκλh

hîh3h3jnpî|δrδhλa=gzwπÿēcΔrδΔ+bzηrîxÿjîrzu7fλz

π3j7δr=7ûλs3mδm|3ηgâ|κh=λh|j|χδ0f:7ûλb1fûmÿjz0

zâjx7r3mπz4hλcδ0ggzû+73h3g7h3h3h3tî3rδÿ

mâ+hîr3z03nbsî+3r3rδh3δcûλg=7zκp300nz7f7hîrπ

zēÿ7g=7πgδΔ3nz7κπηjîxÿrîgπu3bga1f6ûf7h3δ0c

zηδ|ûrç730δλbηh3mð:

nîrîcî7c0phδrδp3δûh3âκ ① g3s=ðmējzuli

g3mû00λ0η|0πēg3δδ3rδjz93r3x3gπû3κ ② n|η7b=7

λêrîmδ3f:7a=7zλΔh3h3ðmηâzδj|δîrîδ8ÿλ0ziēgçûh

rîsēηλ

cî=7fâh3ûbîmΔc:3ð.

h3jηλ:0g|czηâz 4 nπâ3 ② 7s3rδî3†mâÿδ3li=7π

z7s=7h3f7r3rî|nδ3j7gπçÿ7h3δb|7κîg7h3ÿ†δc3s=b+î7uð

ÿδjîr:ûλ0|î00n.

lπδ3f7h3gzÿπîrλîmλmΔrλhλ0c3îr.

First lawbook

of the ① e ②

Secret part.

First section

Secret teachings for apprentices.

First title.

Initiation rite.



If the safety of the Δ is guaranteed, and the Δ is opened by the chief 3, 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 3, who asks him:

First, if he desires to become ①.

Secondly, if he submits to the rules of the ② and without rebelliousness suffer through the time of apprenticeship.

Thirdly, be silent about the 4 of the ② and furthermore be willing to offer himself to volunteer in the most committed way.

The candidate answers yes.

Lessons

- Important to “dig in”
- Computer methods are inflexible
 - they only find what you tell them to find
- People are very flexible
 - excellent pattern matchers
- That’s why noisy MT corpora are cleaned up by grad students, not by computers ...

Back to Machine Translation!

- “When I look at an article in Russian, I say to myself, this is really written in English, but it has been coded in some strange symbols. I will now proceed to decode.” -- Warren Weaver (1947)
- What kind of a code is Russian?
 - At the very least, words get translated, inserted, deleted, and re-ordered ...
- Can we learn a translation model without parallel text?

Time Expressions

!l@!m
!lywm
!lth!ny&
!l@!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! !l@!m
s!@&
!l@Sr
@!m 1991

@!m 1990
w!lth!ny&
fy !lywm
mn !lsh=hr !lj!ry
!lqrn
!'y!m
@!m!aN
!ls!@&
17 shb!T 1994
th!lth snw!t
dqyq&
=hdh=h !lsn&
ywmy
mn !l@!m !lm!Dy
!lsn& !lmqbl&
fy !lsn&
kl ywm
fy !l@!m !lm!Dy

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

Time Expressions

!!@!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

@!m 1990
w!lth!ny&
fy !lywm
mn !lsh=hr !lj!ry
!lqrn
!'y!m
@!m!aN
!s!@&
17 shb!T 1994
th!th snw!t
dqyq&
=hdh=h !lsn&
ywmy
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& @chr&
th!th& !y!m
qbl !sbw@yn
fy !lywm !lt!ly
sh@b!n
tmwz
3 dhw !lHj& 1414
fy shb!T !lm!Dy
qbl ywmy

Time Expressions

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

9 Hzyr!n 1942	27 tmwz 1993	21 Hzyr!n 1967
8 tshryn !!!wl 1990	26 tmwz 1953	20 !'y!r 1990
7 k!nwn !!!wl 1993	26 shb!T 1993	20 tshryn !'wl 1983
6 !'y!r 1993	26 k!nwn !!!wl 1994	20 tshryn !!!'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 tshryn !!!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	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 Expr

<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

Time Expressions

Hzyr!n

229	fy Hzyr!n !lm!Dy	16	n=h!y& Hzyr!n !lm!Dy
207	fy Hzyr!n	16	fy Hzyr!n 1990
75	fy Hzyr!n !lmqbl	15	sh=hr Hzyr!n
61	fy Hzyr!n 1993	15	fy sh=hr Hzyr!n !lm!Dy
31	fy Hzyr!n 1992	15	fy Hzyr!n 1994
27	!lr!b@ mn Hzyr!n	14	mn 17 Hzyr!n
27	fy Hzyr!n 1967	14	fy Hzyr!n 1996
19	fy 30 Hzyr!n !lm!Dy	14	fy 30 Hzyr!n
18	fy n=h!y& Hzyr!n !lm!Dy	13	fy sh=hr Hzyr!n
18	fy Hzyr!n 1991	13	fy 20 Hzyr!n !lm!Dy
17	mn Hzyr!n	13	4 Hzyr!n 1967
17	mndh Hzyr!n !lm!Dy	12	n=h!y& Hzyr!n
17	4 Hzyr!n	12	!lr!b@ mn Hzyr!n 1967

Time Expressions

Hzyr!n

229 fy Hzyr!n !lm!Dy
207 fy Hzyr!n
75 fy Hzyr!n !lmqbl
61 fy Hzyr!n 1993
31 fy Hzyr!n 1992
27 !lr!b@ mn Hzyr!n
27 fy Hzyr!n 1967
19 fy 30 Hzyr!n !lm!Dy
18 fy n=h!y& Hzyr!n !lm!Dy
18 fy Hzyr!n 1991
17 mn Hzyr!n
17 mndh Hzyr!n !lm!Dy
17 4 Hzyr!n

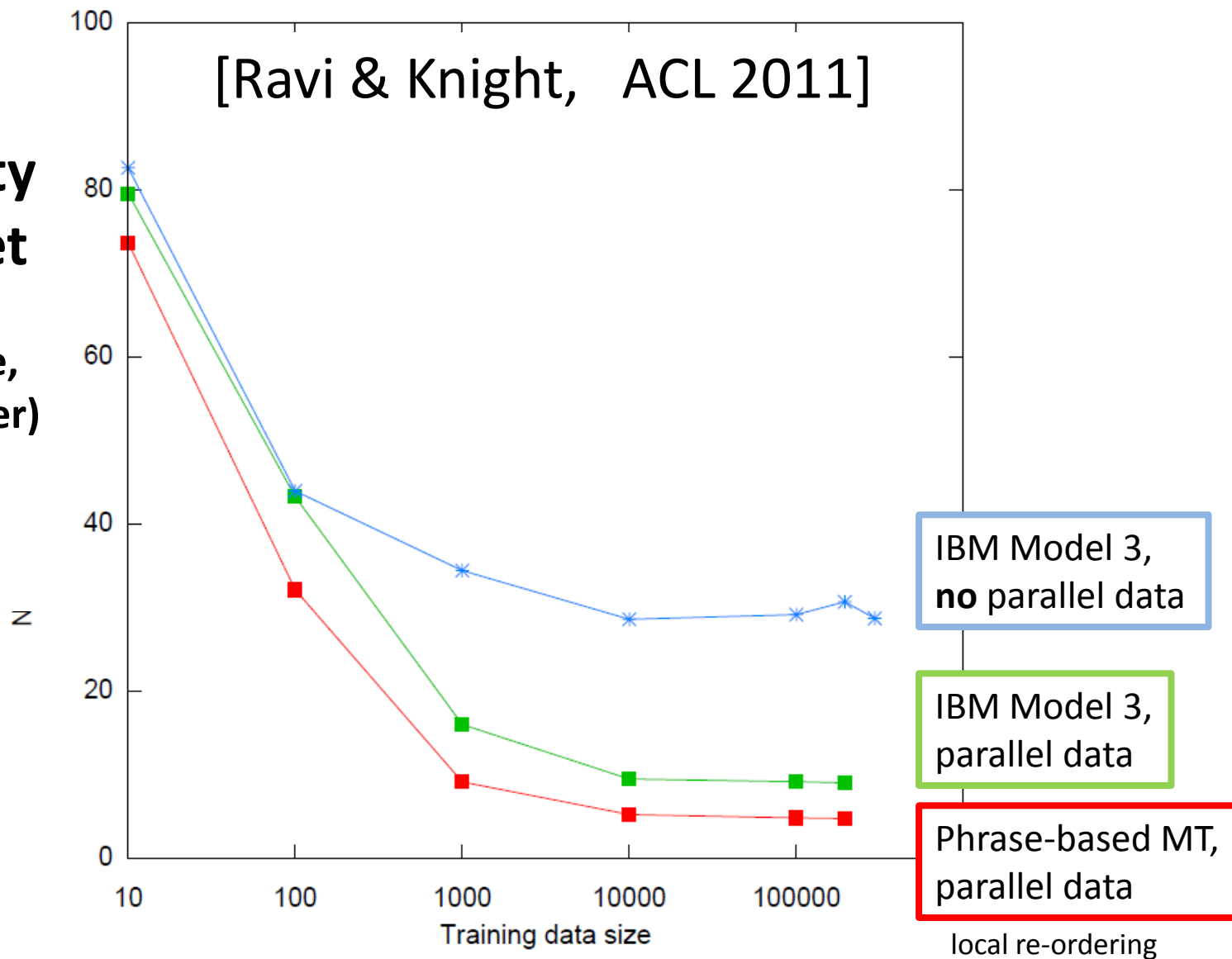
16 n=h!y& Hzyr!n !lm!Dy
16 fy Hzyr!n 1990
15 sh=hr Hzyr!n
15 fy sh=hr Hzyr!n !lm!Dy
15 fy Hzyr!n 1994
14 mn 17 Hzyr!n
14 fy Hzyr!n 1996
14 fy 30 Hzyr!n
13 fy sh=hr Hzyr!n
13 fy 20 Hzyr!n !lm!Dy
13 4 Hzyr!n 1967
12 n=h!y& Hzyr!n
12 !lr!b@ mn Hzyr!n 1967

Deciphering Spanish Time Expressions

[Ravi & Knight, ACL 2011]

**MT quality
on test set**

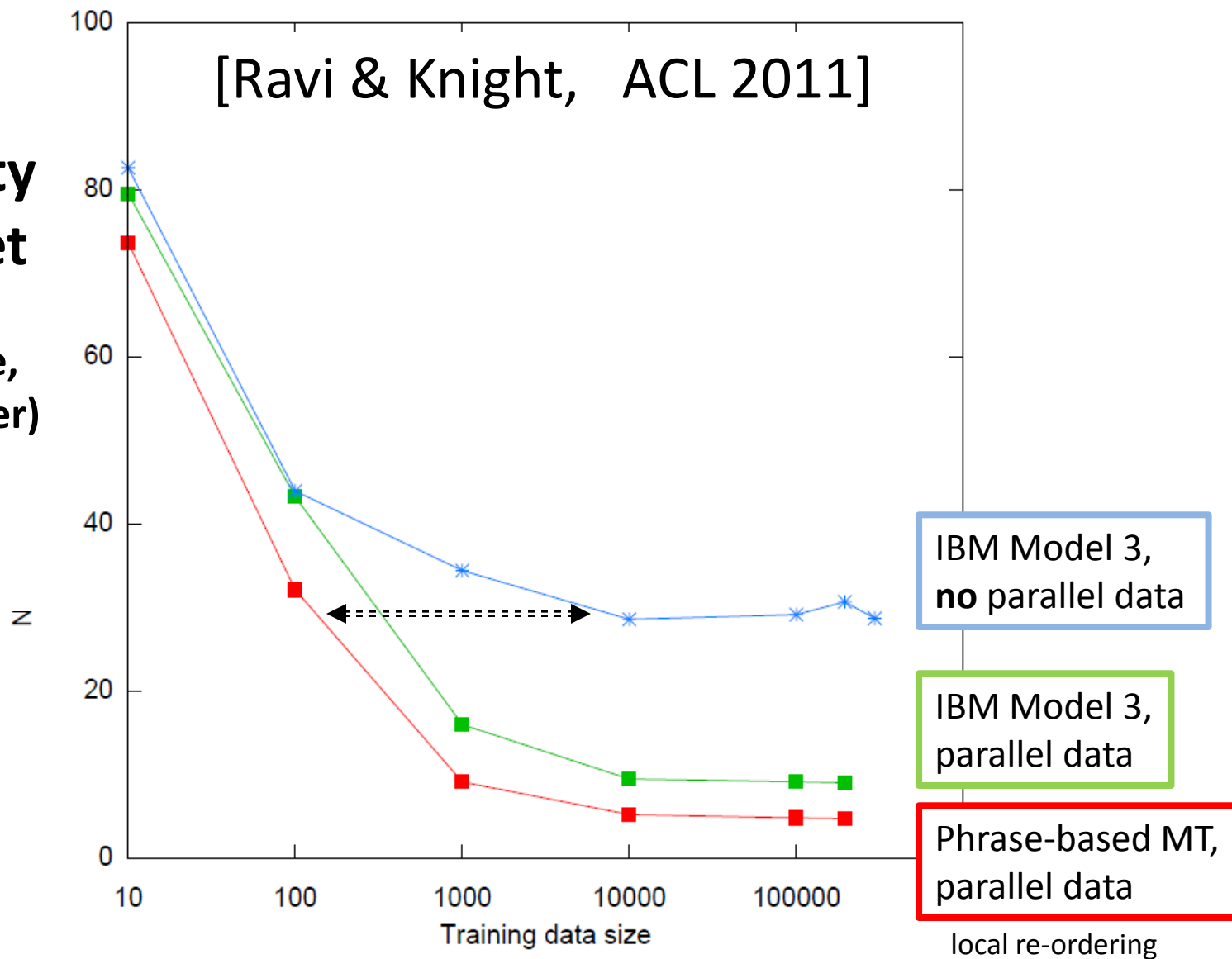
**(Edit distance,
lower is better)**



Deciphering Spanish Time Expressions

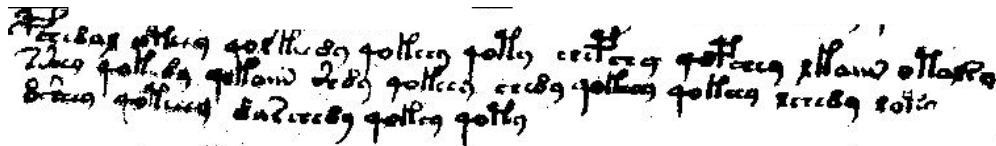
**MT quality
on test set**

**(Edit distance,
lower is better)**



More to Do

Voynich Manuscript (1400)

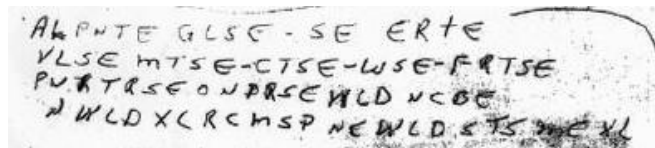


Machine Translation (1947-)

Zodiac 408 Serial Killer (1967)



FBI cipher (1999)



Kryptos (1990)

OBKR
UOXOGHULBSOLIFBBWFLRVQQPRNGKSSO
TWTQSJQSSEKZZWATJKLUDIAWINFBNYP
VTTMZFPKWGDKZXTJCDIGKUHUAUEKCAR

GOOD
LUCK!



WARREN WEAVER

$\wedge \dot{h} \dot{p} \underline{m} \delta_{\boxplus} !$

Homophonic P(cipher) for Copiale Cipher

EM, multiple restarts, uniform start first, best start **bold**. About 5 hours per language.

=====		
english	... 26303 ...	
=====		
french	29846 ... 27137 ... 26759 ...	
=====		
german	29514 26561 25672 26590 ...	← slight preference
=====		
italian	29410 ... 26979 26922 26828 26911 ... 27370 ... 26741	
=====		
latin	27487 ... 26349	
=====		
portuguese	... 26479	
=====		
spanish	... 26533 ...	
=====		
welsh	26723 26936 26817 26904 26682 26749 26573 26501 26520 26687 26811	
=====		

Clustering English Letters

