

Language Translation and Code-Breaking

Kevin Knight
Information Sciences Institute
University of Southern California

joint work with

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

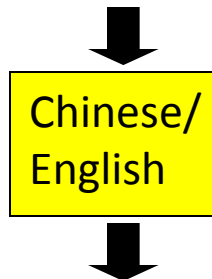
Machine Translation

美国关岛国际机场及其办公室均接获一名自称沙地阿拉伯富商拉登等发出的电子邮件，威胁将会向机场等公众地方发动生化袭击後，关岛经保持高度戒备。

Kowane mutum na da hakkin ya sami yancin yin tunani da na sanin yakamata da na bin addini; saboda haka yana da yancin sake addini ko ra'ayin da ya bada gaskiya gare shi, da kuma yancin nuna addininsa ko ra'ayinsa, shi daya ko a cikin taro kuma a fili ko a boye ta hanyar koyarwa ko yin ibada, ko bauta wa abin da ya bada gaskiya gare shi da yin abubuwan da abin da yake bauta wa din ya nuna masa.

Machine Translation

美国关岛国际机场及其办公室均接获一名自称沙地阿拉伯富商拉登等发出的电子邮件，威胁将会向机场等公众地方发动生化袭击後，关岛经保持高度戒备。



The U.S. island of Guam is maintaining a high state of alert after the Guam airport and its offices both received an e-mail from someone calling himself the Saudi Arabian Osama bin Laden and threatening a biological/chemical attack against public places such as the airport.

Kowane mutum na da hakkin ya sami yancin yin tunani da na sanin yakamata da na bin addini; saboda haka yana da yancin sake addini ko ra'ayin da ya bada gaskiya gare shi, da kuma yancin nuna addininsa ko ra'ayinsa, shi daya ko a cikin taro kuma a fili ko a boye ta hanyar koyarwa ko yin ibada, ko bauta wa abin da ya bada gaskiya gare shi da yin abubuwan da abin da yake bauta wa din ya nuna masa.



Everyone has the right to freedom of thought, conscience and religion; this right includes freedom to change his religion or belief, and freedom, either alone or in community with others and in public or private, to manifest his religion or belief in teaching, practice, worship and observance.

Statistical Machine Translation

“When I look at an article in Russian, I say: this is really written in English, but it has been coded in some strange symbols. I will now proceed to decode.” -- Warren Weaver (1947)

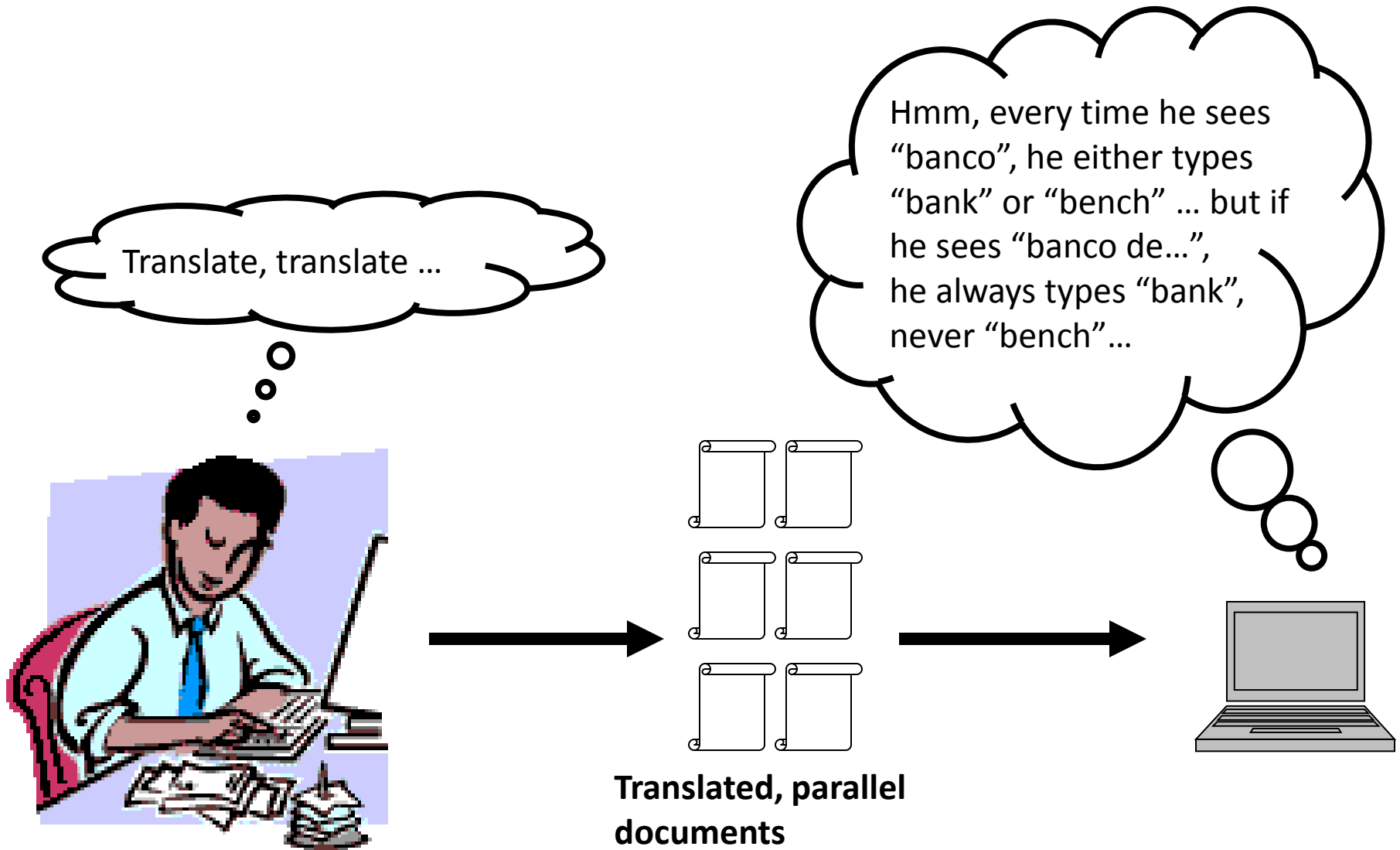


OUR HERO

Weaver saw a colleague decoding intercepts into Turkish, without “knowing” Turkish.

... maybe a computer could translate into English without “knowing” English?

Statistical Machine Translation



Parallel Corpus

12 English sentences in English and Spanish.

1a. Garcia and associates .
1b. Garcia y asociados .

7a. the clients and the associates are enemies .
7b. los clients y los asociados son enemigos .

2a. Carlos Garcia has three associates .
2b. Carlos Garcia tiene tres asociados .

8a. the company has three groups .
8b. la empresa tiene tres grupos .

3a. his associates are not strong .
3b. sus asociados no son fuertes .

9a. its groups are in Europe .
9b. sus grupos estan en Europa .

4a. Garcia has a company also .
4b. Garcia tambien tiene una empresa .

10a. the modern groups sell strong pharmaceuticals .
10b. los grupos modernos venden medicinas fuertes .

5a. its clients are angry .
5b. sus clientes estan enfadados .

11a. the groups do not sell zenzanine .
11b. los grupos no venden zanzanina .

6a. the associates are also angry .
6b. los asociados tambien estan enfadados .

12a. the small groups are not modern .
12b. los grupos pequenos no son modernos .

Parallel Corpus

12 English sentences in Centauri and Arcturan.

| | |
|-----------------------------------------|--------------------------------------------------|
| 1a. ok-voon ororok sprok . | 7a. lalok farok ororok lalok sprok izok enemok . |
| 1b. at-voon bichat dat . | 7b. wat jjat bichat wat dat vat eneat . |
| 2a. ok-drubel ok-voon anak plok sprok . | 8a. lalok brok anak plok nok . |
| 2b. at-drubel at-voon pippat rrat dat . | 8b. iat lat pippat rrat nnat . |
| 3a. erok sprok izok hihok ghrok . | 9a. wiwok nok izok kantok ok-yurp . |
| 3b. totat dat arrat vat hilat . | 9b. totat nnat quat oloat at-yurp . |
| 4a. ok-voon anak drok brok jok . | 10a. lalok mok nok yorok ghrok klok . |
| 4b. at-voon krat pippat sat lat . | 10b. wat nnat gat mat bat hilat . |
| 5a. wiwok farok izok stok . | 11a. lalok nok crrok hihok yorok zanzanok . |
| 5b. totat jjat quat cat . | 11b. wat nnat arrat mat zanzanat . |
| 6a. lalok sprok izok jok stok . | 12a. lalok rarok nok izok hihok mok . |
| 6b. wat dat krat quat cat . | 12b. wat nnat forat arrat vat gat . |

Centauri/Arcturan

Your assignment, translate this to Arcturan: farok crrrok hihok yorok klok kantok ok-yurp

| | |
|-----------------------------------------|--------------------------------------------------|
| 1a. ok-voon ororok sprok . | 7a. lalok farok ororok lalok sprok izok enemok . |
| 1b. at-voon bichat dat . | 7b. wat jjat bichat wat dat vat eneat . |
| 2a. ok-drubel ok-voon anak plok sprok . | 8a. lalok brok anak plok nok . |
| 2b. at-drubel at-voon pippat rrat dat . | 8b. iat lat pippat rrat nnat . |
| 3a. erok sprok izok hihok ghrok . | 9a. wiwok nok izok kantok ok-yurp . |
| 3b. totat dat arrat vat hilat . | 9b. totat nnat quat oloat at-yurp . |
| 4a. ok-voon anak drok brok jok . | 10a. lalok mok nok yorok ghrok klok . |
| 4b. at-voon krat pippat sat lat . | 10b. wat nnat gat mat bat hilat . |
| 5a. wiwok farok izok stok . | 11a. lalok nok crrrok hihok yorok zanzanok . |
| 5b. totat jjat quat cat . | 11b. wat nnat arrat mat zanzanat . |
| 6a. lalok sprok izok jok stok . | 12a. lalok rarok nok izok hihok mok . |
| 6b. wat dat krat quat cat . | 12b. wat nnat forat arrat vat gat . |

Centauri/Arcturan

Your assignment, translate this to Arcturan: **farok** crrrok hihok yorok klok kantok ok-yurp

| | |
|-----------------------------------------|---------------------------------------------------------|
| 1a. ok-voon ororok sprok . | 7a. lalok farok ororok lalok sprok izok enemok . |
| 1b. at-voon bichat dat . | 7b. wat jjat bichat wat dat vat eneat . |
| 2a. ok-drubel ok-voon anak plok sprok . | 8a. lalok brok anak plok nok . |
| 2b. at-drubel at-voon pippat rrat dat . | 8b. iat lat pippat rrat nnat . |
| 3a. erok sprok izok hihok ghrok . | 9a. wiwok nok izok kantok ok-yurp . |
| 3b. totat dat arrat vat hilat . | 9b. totat nnat quat oloat at-yurp . |
| 4a. ok-voon anak drok brok jok . | 10a. lalok mok nok yorok ghrok klok . |
| 4b. at-voon krat pippat sat lat . | 10b. wat nnat gat mat bat hilat . |
| 5a. wiwok farok izok stok . | 11a. lalok nok crrrok hihok yorok zanzanak . |
| 5b. totat jjat quat cat . | 11b. wat nnat arrat mat zanzanat . |
| 6a. lalok sprok izok jok stok . | 12a. lalok rarok nok izok hihok mok . |
| 6b. wat dat krat quat cat . | 12b. wat nnat forat arrat vat gat . |

Centauri/Arcturan

Your assignment, translate this to Arcturan: farok **crrok** hihok yorok klok kantok ok-yurp

| | |
|-----------------------------------------|---------------------------------------------------------|
| 1a. ok-voon ororok sprok . | 7a. lalok farok ororok lalok sprok izok enemok . |
| 1b. at-voon bichat dat . | 7b. wat jjat bichat wat dat vat eneath . |
| 2a. ok-drubel ok-voon anak plok sprok . | 8a. lalok brok anak plok nok . |
| 2b. at-drubel at-voon pippat rrat dat . | 8b. iat lat pippat rrat nnat . |
| 3a. erok sprok izok hihok ghrok . | 9a. wiwok nok izok kantok ok-yurp . |
| 3b. totat dat arrat vat hihat . | 9b. totat nnat quat oloat at-yurp . |
| 4a. ok-voon anak drok brok jok . | 10a. lalok mok nok yorok ghrok klok . |
| 4b. at-voon krat pippat sat lat . | 10b. wat nnat gat mat bat hihat . |
| 5a. wiwok farok izok stok . | 11a. lalok nok crrok hihok yorok zanzanak . |
| 5b. totat jjat quat cat . | 11b. wat nnat arrat mat zanzanat . |
| 6a. lalok sprok izok jok stok . | 12a. lalok rarok nok izok hihok mok . |
| 6b. wat dat krat quat cat . | 12b. wat nnat forat arrat vat gat . |

Centauri/Arcturan

Your assignment, translate this to Arcturan: **farok** **crrok** hihok yorok klok kantok ok-yurp

| | |
|-----------------------------------------|---------------------------------------------------------|
| 1a. ok-voon ororok sprok . | 7a. lalok farok ororok lalok sprok izok enemok . |
| 1b. at-voon bichat dat . | / 7b. wat jjat bichat wat dat vat eneat . |
| 2a. ok-drubel ok-voon anak plok sprok . | 8a. lalok brok anak plok nok . |
| 2b. at-drubel at-voon pippat rrat dat . | 8b. iat lat pippat rrat nnat . |
| 3a. erok sprok izok hihok ghrok . | 9a. wiwok nok izok kantok ok-yurp . |
| 3b. totat dat arrat vat hilat . | 9b. totat nnat quat oloat at-yurp . |
| 4a. ok-voon anak drok brok jok . | 10a. lalok mok nok yorok ghrok klok . |
| 4b. at-voon krat pippat sat lat . | 10b. wat nnat gat mat bat hilat . |
| 5a. wiwok farok izok stok . | 11a. lalok nok crrok hihok yorok zanzanok . |
| / | ??? |
| 5b. totat jjat quat cat . | 11b. wat nnat arrat mat zanzanat . |
| 6a. lalok sprok izok jok stok . | 12a. lalok rarok nok izok hihok mok . |
| 6b. wat dat krat quat cat . | 12b. wat nnat forat arrat vat gat . |

Centauri/Arcturan

Your assignment, translate this to Arcturan: **farok** crrok **hihok** yorok klok kantok ok-yurp

| | |
|------------------------------------------|----------------------------------------------------|
| 1a. ok-voon ororok sprok . | 7a. lalok farok ororok lalok sprok izok enemok . |
| 1b. at-voon bichat dat . | / |
| | 7b. wat jjat bichat wat dat vat eneat . |
| 2a. ok-drubel ok-voon anak plok sprok . | 8a. lalok brok anak plok nok . |
| 2b. at-drubel at-voon pippat rrat dat . | 8b. iat lat pippat rrat nnat . |
| 3a. erok sprok izok hihok ghrok . | 9a. wiwok nok izok kantok ok-yurp . |
| 3b. totat dat arrat vat hilat . | 9b. totat nnat quat oloat at-yurp . |
| 4a. ok-voon anak drok brok jok . | 10a. lalok mok nok yorok ghrok klok . |
| 4b. at-voon krat pippat sat lat . | 10b. wat nnat gat mat bat hilat . |
| 5a. wiwok farok izok stok . | 11a. lalok nok crrok hihok yorok zanzanok . |
| / | 11b. wat nnat arrat mat zanzanat . |
| 5b. totat jjat quat cat . | 12a. lalok rarok nok izok hihok mok . |
| 6a. lalok sprok izok jok stok . | |
| 6b. wat dat krat quat cat . | 12b. wat nnat forat arrat vat gat . |

Centauri/Arcturan

Your assignment, translate this to Arcturan: **farok** crrok **hihok** **yorok** klok kantok ok-yurp

| | |
|------------------------------------------|-----------------------------------------------------------|
| 1a. ok-voon ororok sprok . | 7a. lalok farok ororok lalok sprok izok enemok . |
| 1b. at-voon bichat dat . | 7b. wat jjat bichat wat dat vat eneat . |
| 2a. ok-drubel ok-voon anak plok sprok . | 8a. lalok brok anak plok nok . |
| 2b. at-drubel at-voon pippat rrat dat . | 8b. iat lat pippat rrat nnat . |
| 3a. erok sprok izok hihok ghrok . | 9a. wiwok nok izok kantok ok-yurp . |
| 3b. totat dat arrat vat hilat . | 9b. totat nnat quat oloat at-yurp . |
| 4a. ok-voon anak drok brok jok . | 10a. lalok mok nok yorok ghrok klok . |
| 4b. at-voon krat pippat sat lat . | 10b. wat nnat gat mat bat hilat . |
| 5a. wiwok farok izok stok . | 11a. lalok nok crrok hihok yorok zanzanok . |
| 5b. totat jjat quat cat . | 11b. wat nnat arrat mat zanzanat . |
| 6a. lalok sprok izok jok stok . | 12a. lalok rarok nok izok hihok mok . |
| 6b. wat dat krat quat cat . | 12b. wat nnat forat arrat vat gat . |

Centauri/Arcturan

Your assignment, translate this to Arcturan: **farok** crrok **hihok** **yorok** **clok** kantok ok-yurp

| | |
|-----------------------------------------|--------------------------------------------------|
| 1a. ok-voon ororok sprok . | 7a. lalok farok ororok lalok sprok izok enemok . |
| 1b. at-voon bichat dat . | 7b. wat jjat bichat wat dat vat eneat . |
| 2a. ok-drubel ok-voon anak plok sprok . | 8a. lalok brok anak plok nok . |
| 2b. at-drubel at-voon pippat rrat dat . | 8b. iat lat pippat rrat nnat . |
| 3a. erok sprok izok hihok ghrok . | 9a. wiwok nok izok kantok ok-yurp . |
| 3b. totat dat arrat vat hilat . | 9b. totat nnat quat oloat at-yurp . |
| 4a. ok-voon anak drok brok jok . | 10a. lalok mok nok yorok ghrok clok . |
| 4b. at-voon krat pippat sat lat . | 10b. wat nnat gat mat bat hilat . |
| 5a. wiwok farok izok stok . | 11a. lalok nok crrok hihok yorok zanzanok . |
| 5b. totat jjat quat cat . | 11b. wat nnat arrat mat zanzanat . |
| 6a. lalok sprok izok jok stok . | 12a. lalok rarok nok izok hihok mok . |
| 6b. wat dat krat quat cat . | 12b. wat nnat forat arrat vat gat . |

Centauri/Arcturan

Your assignment, translate this to Arcturan: **farok** crrok **hihok** **yorok** **clok** kantok ok-yurp

| | |
|-----------------------------------------|--------------------------------------------------|
| 1a. ok-voon ororok sprok . | 7a. lalok farok ororok lalok sprok izok enemok . |
| 1b. at-voon bichat dat . | 7b. wat jjat bichat wat dat vat eneat . |
| 2a. ok-drubel ok-voon anak plok sprok . | 8a. lalok brok anak plok nok . |
| 2b. at-drubel at-voon pippat rrat dat . | 8b. iat lat pippat rrat nnat . |
| 3a. erok sprok izok hihok ghrok . | 9a. wiwok nok izok kantok ok-yurp . |
| 3b. totat dat arrat vat hilat . | 9b. totat nnat quat oloat at-yurp . |
| 4a. ok-voon anak drok brok jok . | 10a. lalok mok nok yorok ghrok clok . |
| 4b. at-voon krat pippat sat lat . | 10b. wat nnat gat mat bat hilat . |
| 5a. wiwok farok izok stok . | 11a. lalok nok crrok hihok yorok zanzanok . |
| 5b. totat jjat quat cat . | 11b. wat nnat arrat mat zanzanat . |
| 6a. lalok sprok izok jok stok . | 12a. lalok rarok nok izok hihok mok . |
| 6b. wat dat krat quat cat . | 12b. wat nnat forat arrat vat gat . |

Centauri/Arcturan

Your assignment, translate this to Arcturan: **farok** crrok **hihok** **yorok** **clock** kantok ok-yurp

| | |
|------------------------------------------|---------------------------------------------------------|
| 1a. ok-voon ororok sprok . | 7a. lalok farok ororok lalok sprok izok enemok . / |
| 1b. at-voon bichat dat . | 7b. wat jjat bichat wat dat vat eneat . |
| 2a. ok-drubel ok-voon anak plok sprok . | 8a. lalok brok anak plok nok . / |
| 2b. at-drubel at-voon pippat rrat dat . | 8b. iat lat pippat rrat nnat . |
| 3a. erok sprok izok hihok ghrok . / / | 9a. wiwok nok izok kantok ok-yurp . |
| 3b. totat dat arrat vat hilat . | 9b. totat nnat quat oloat at-yurp . |
| 4a. ok-voon anak drok brok jok . | 10a. lalok mok nok yorok ghrok clock . / / / |
| 4b. at-voon krat pippat sat lat . | 10b. wat nnat gat mat bat hilat . |
| 5a. wiwok farok izok stok . / | 11a. lalok nok crrok hihok yorok zanzanok . / / / |
| 5b. totat jjat quat cat . | 11b. wat nnat arrat mat zanzanat . |
| 6a. lalok sprok izok jok stok . | 12a. lalok rarok nok izok hihok mok . / / / |
| 6b. wat dat krat quat cat . | 12b. wat nnat forat arrat vat gat . |

Centauri/Arcturan

Your assignment, translate this to Arcturan: **farok** crrok **hihok** **yorok** **clock** kantok ok-yurp

| | |
|------------------------------------------|-------------------------------------------------------------------------|
| 1a. ok-voon ororok sprok . | 7a. lalok farok ororok lalok sprok izok enemok . / |
| 1b. at-voon bichat dat . | 7b. wat jjat bichat wat dat vat eneat . |
| 2a. ok-drubel ok-voon anak plok sprok . | 8a. lalok brok anak plok nok . / |
| 2b. at-drubel at-voon pippat rrat dat . | 8b. iat lat pippat rrat nnat . |
| 3a. erok sprok izok hihok ghrok . / / | 9a. wiwok nok izok kantok ok-yurp . |
| 3b. totat dat arrat vat hilat . | 9b. totat nnat quat oloat at-yurp . |
| 4a. ok-voon anak drok brok jok . | 10a. lalok mok nok yorok ghrok clock . X / |
| 4b. at-voon krat pippat sat lat . | 10b. wat nnat gat mat bat hilat . / process of elimination |
| 5a. wiwok farok izok stok . / | 11a. lalok nok crrok hihok yorok zanzanok . / / |
| 5b. totat jjat quat cat . | 11b. wat nnat arrat mat zanzanat . / / / |
| 6a. lalok sprok izok jok stok . | 12a. lalok rarok nok izok hihok mok . / / / |
| 6b. wat dat krat quat cat . | 12b. wat nnat forat arrat vat gat . |

Statistical Machine Translation

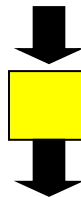
1999

- one page per day, low quality
- limited domains, languages
- no commercial offerings

2011

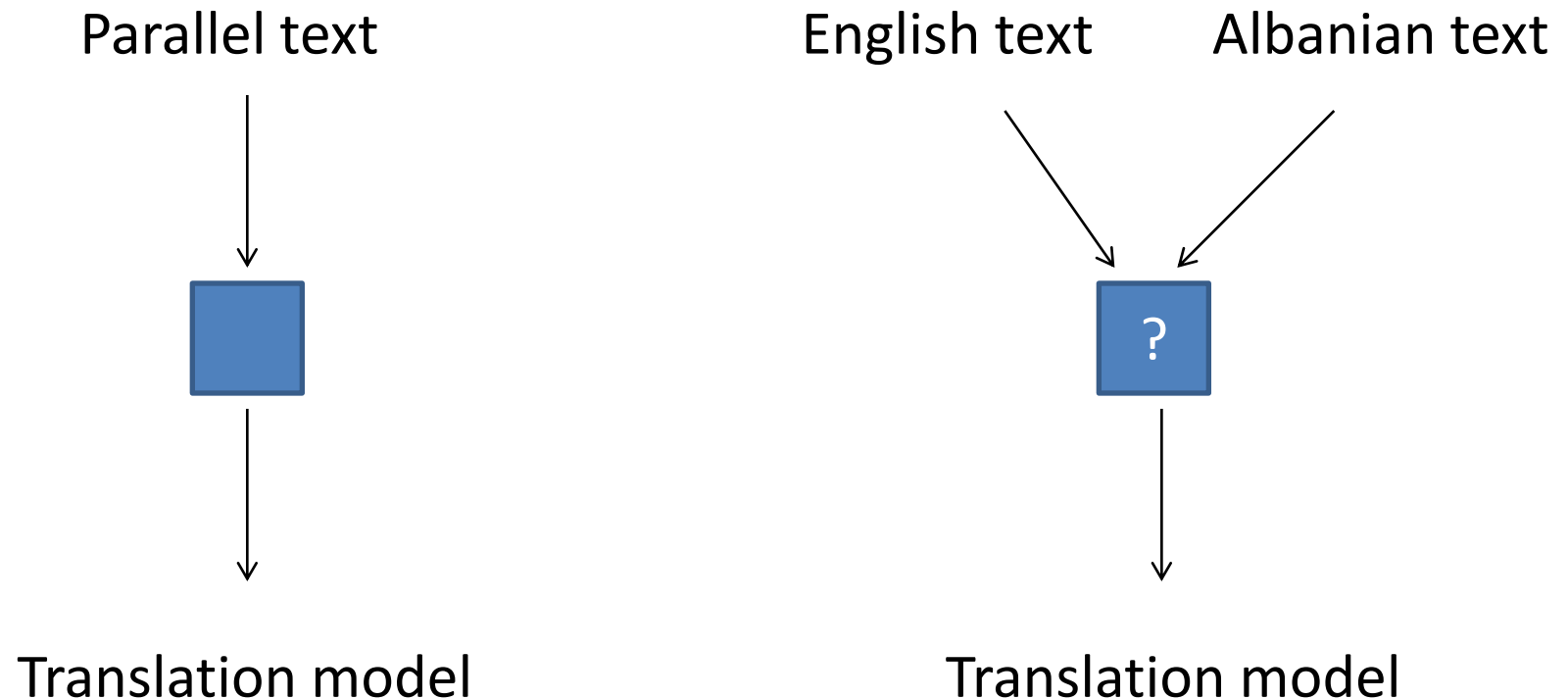
- fast, 50+ languages
- quality much improved
- products for business, intelligence, end users

أعلن الرئيس الصومالي شريف شيخ أحمد أثناء زيارة لمواقع للقوات الحكومية والأفريقية في حي هودن في مقديشو أن الحملة العسكرية الحالية "لن تتوقف حتى تتحرر الصومال من الشباب والقاعدة".



Somali President Sharif Sheikh Ahmed during a visit to sites of government forces and African in the district of Howden in Mogadishu that the current military campaign "will not stop until Somalia is liberated from the youth and al-Qaeda. "

Learn Translation Knowledge from Non-Parallel Text?



German Enigma Machines (1926-45)

Substitution system

$N \rightarrow J$

Substitution table **changes**
with every keystroke:

$NNN \rightarrow JTE$

Flattens out ciphertext
letter distributions.

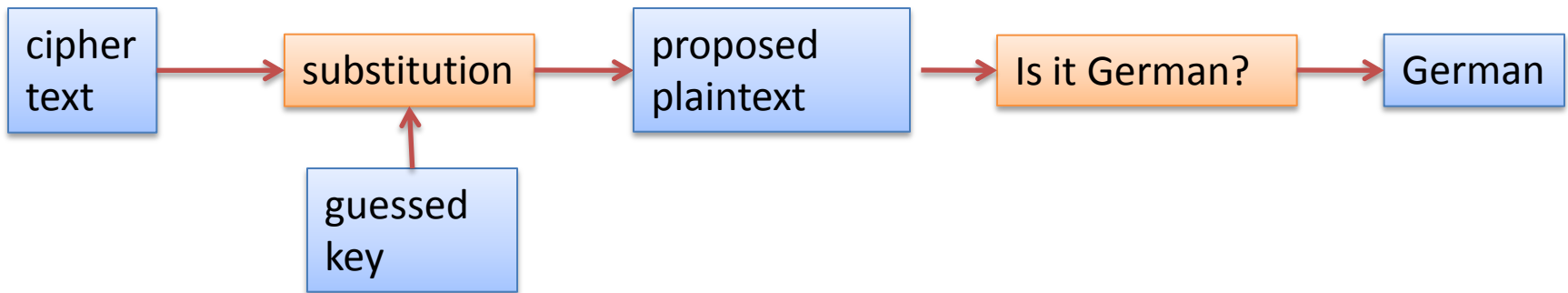


Secret key =
initial rotor
ordering and
settings

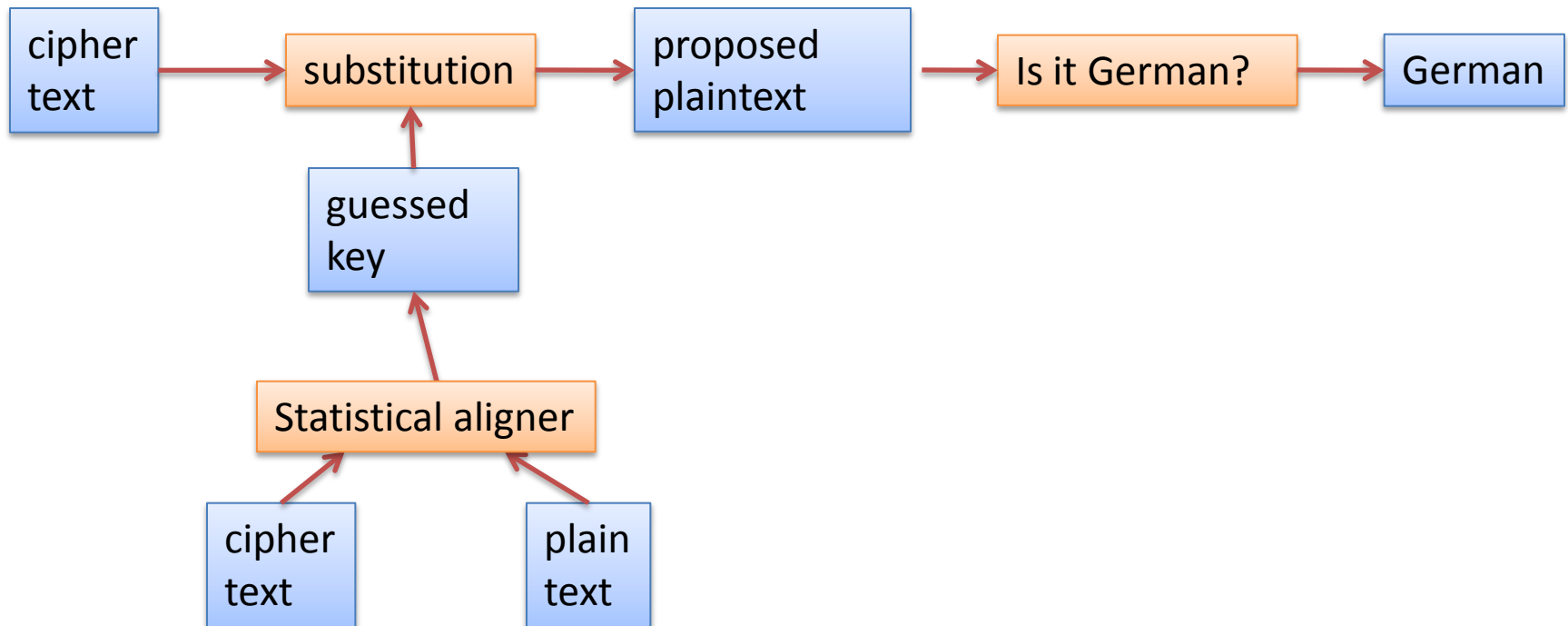
Reversible behavior

$NNN \rightarrow JTE \rightarrow NNN$

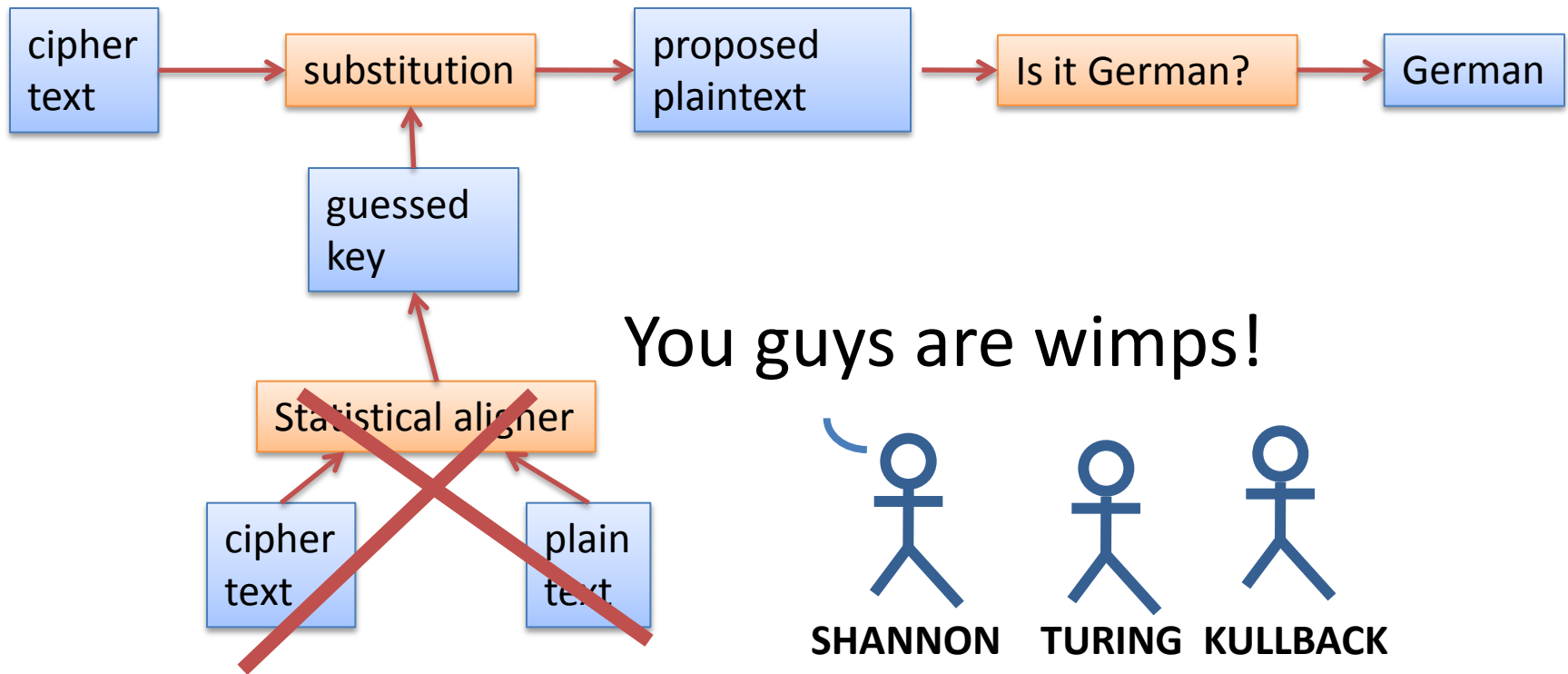
Breaking Enigma



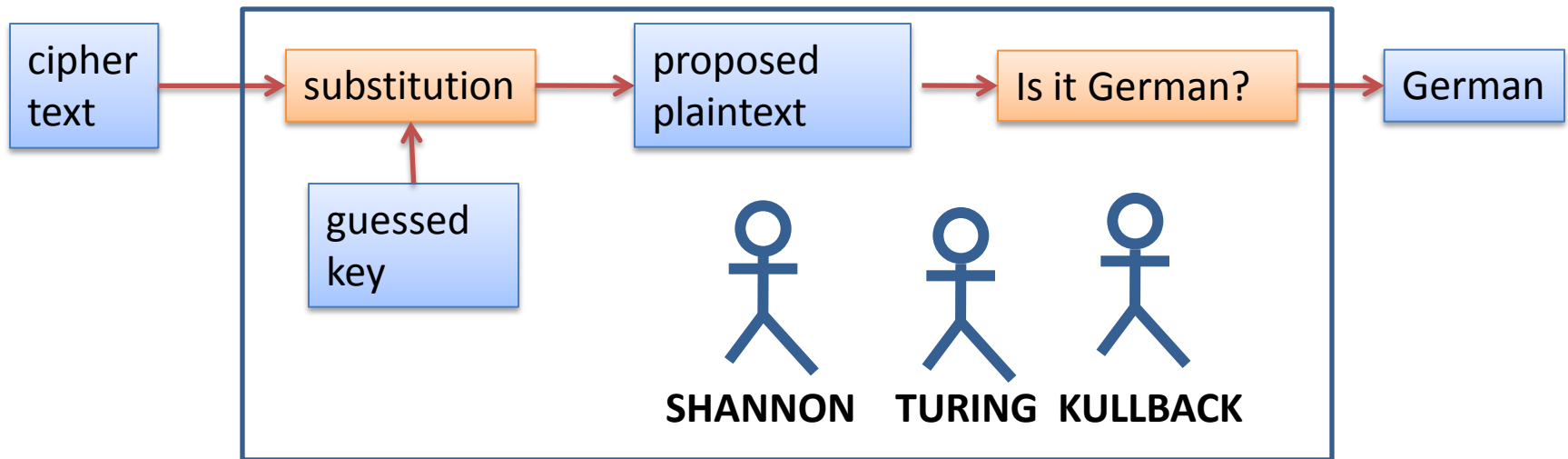
Breaking Enigma



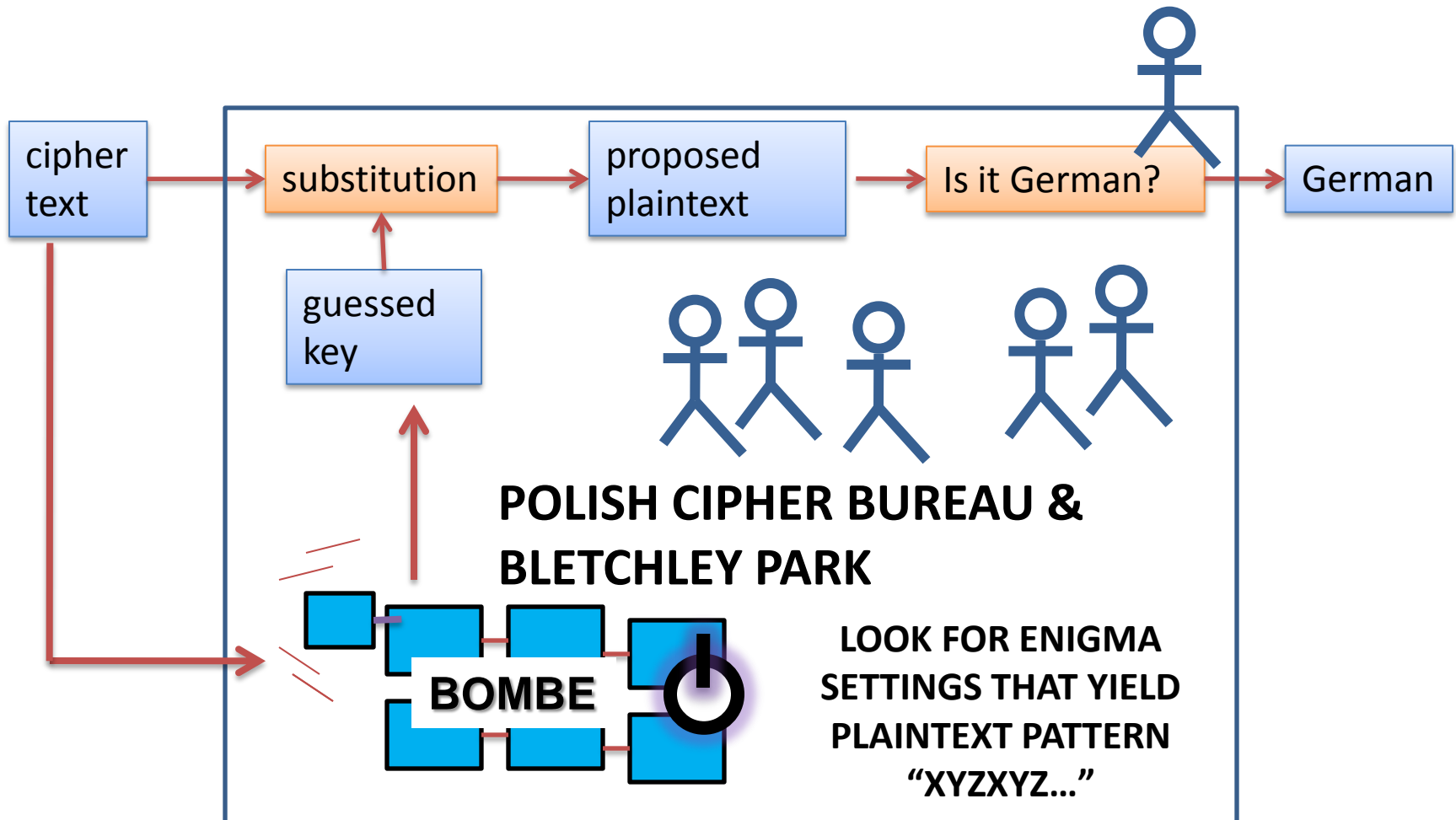
Breaking Enigma



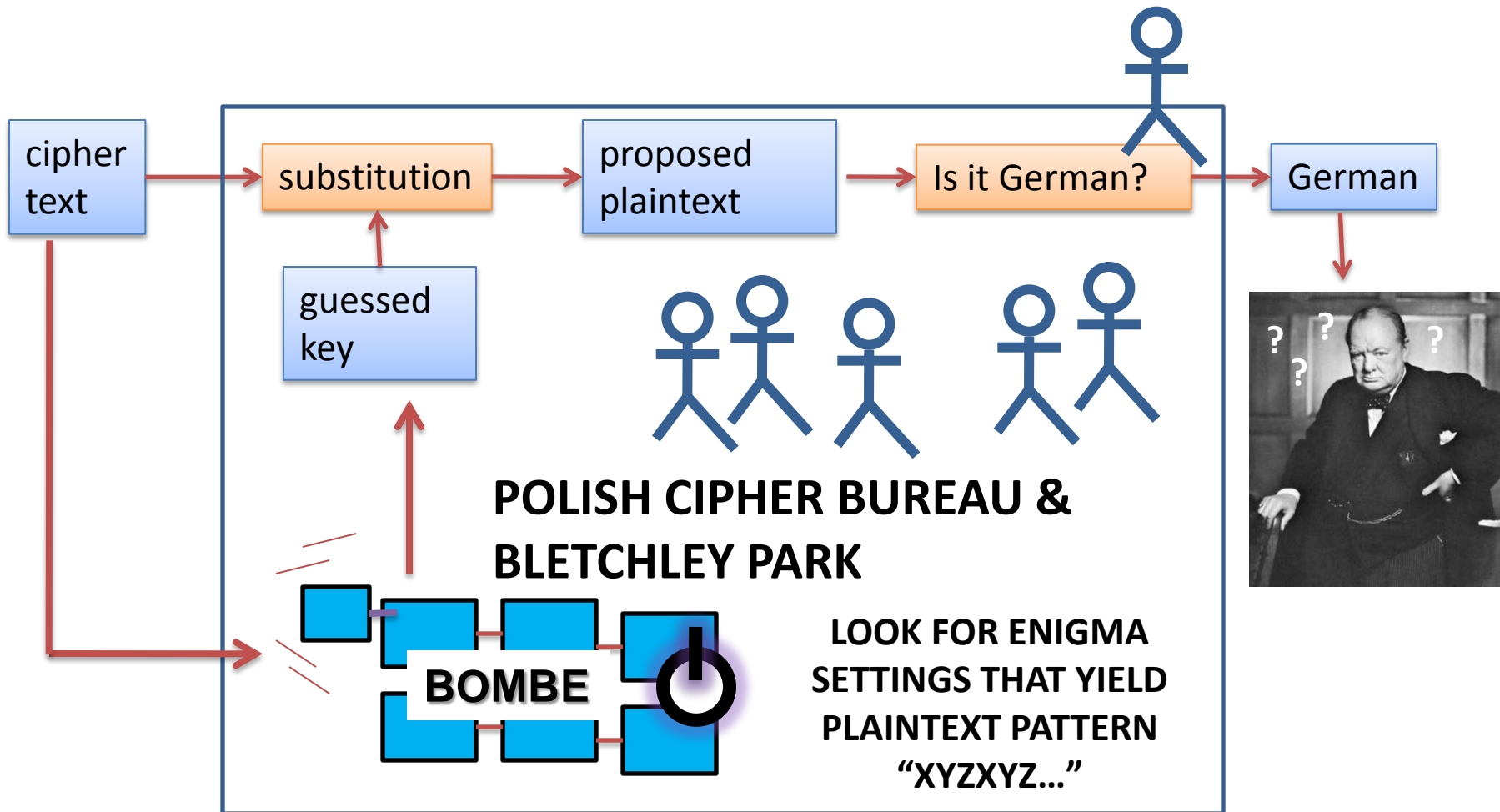
Breaking Enigma



Breaking Enigma



Breaking Enigma

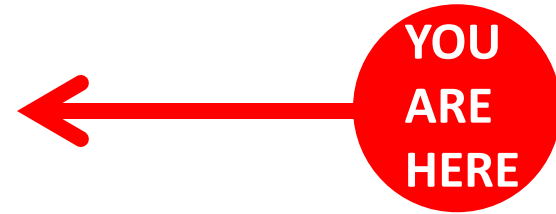


What's the Plan for This Talk?

- Break a series of codes
 - Simple letter substitution
 - Phonetic substitution
 - archaeology
 - transliteration
 - Word substitution
 - Foreign language as cipher
- Bonus
 - Two historical ciphers
 - Final thought on translation and cryptography

What's the Plan for This Talk?

- Break a series of codes
 - Simple letter substitution
 - Phonetic substitution
 - archaeology
 - transliteration
 - Word substitution
 - Foreign language as cipher
- Bonus
 - Two historical ciphers
 - Final thought on translation and cryptography



Letter Substitution Cipher

- Encipherment key:

PLAIN: ABCDEFGHIJKLMNOPQRSTUVWXYZ

CIPHER: PLOKMIJNUHBYGVTFCDXESZQW

- 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 ##### 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 ~~X~~ 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

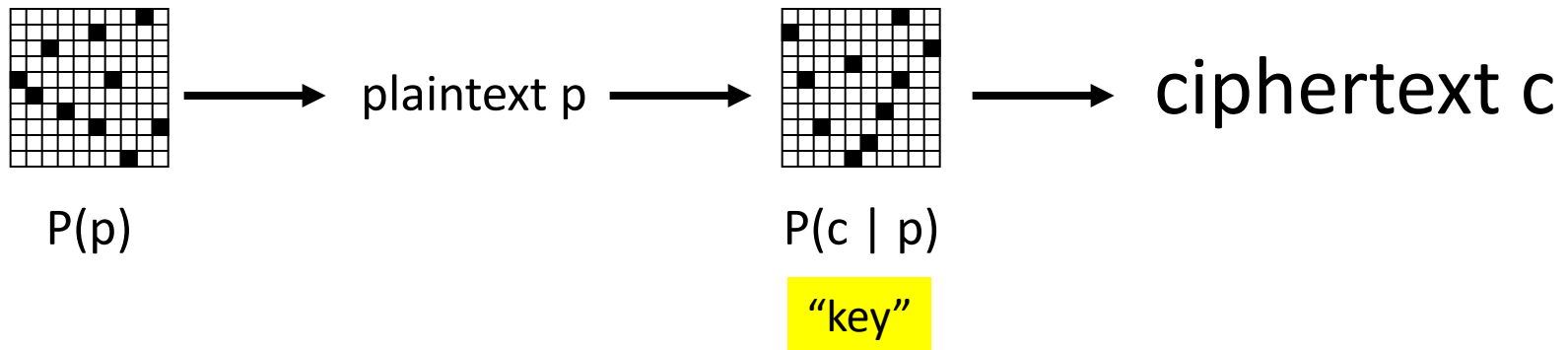
discover a way to repair it
PTBMYEQL K FKH CY LQZKTL TC."

| | |
|---|------------|
| 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 . |

Letter Substitution Cipher

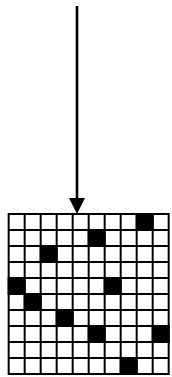
ciphertext c

Letter Substitution Cipher



Letter Substitution Cipher

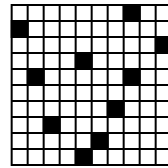
plaintext samples,
unrelated to ciphertext



$P(p)$



plaintext p



$P(c \mid p)$

“key”



ciphertext c

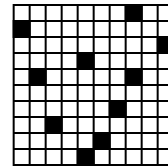
Letter Substitution Cipher

plaintext samples,
unrelated to ciphertext



$P(p)$

plaintext p

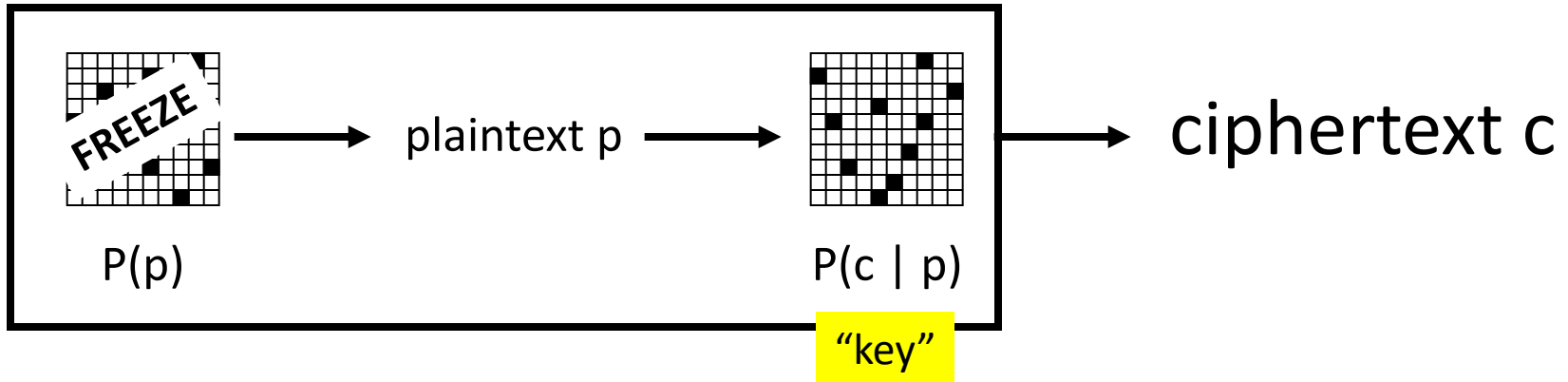


$P(c \mid p)$

"key"

ciphertext c

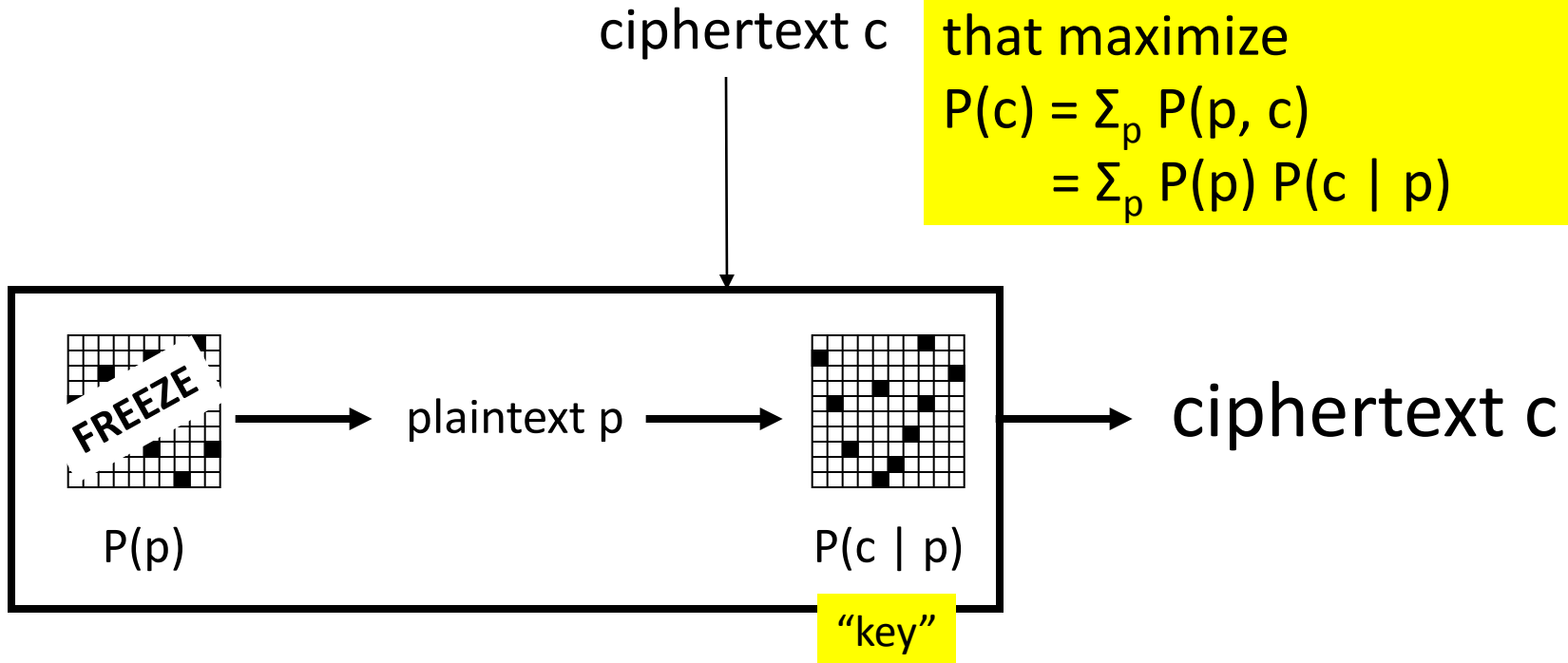
Letter Substitution Cipher



Letter Substitution Cipher

Find substitution-table values that maximize

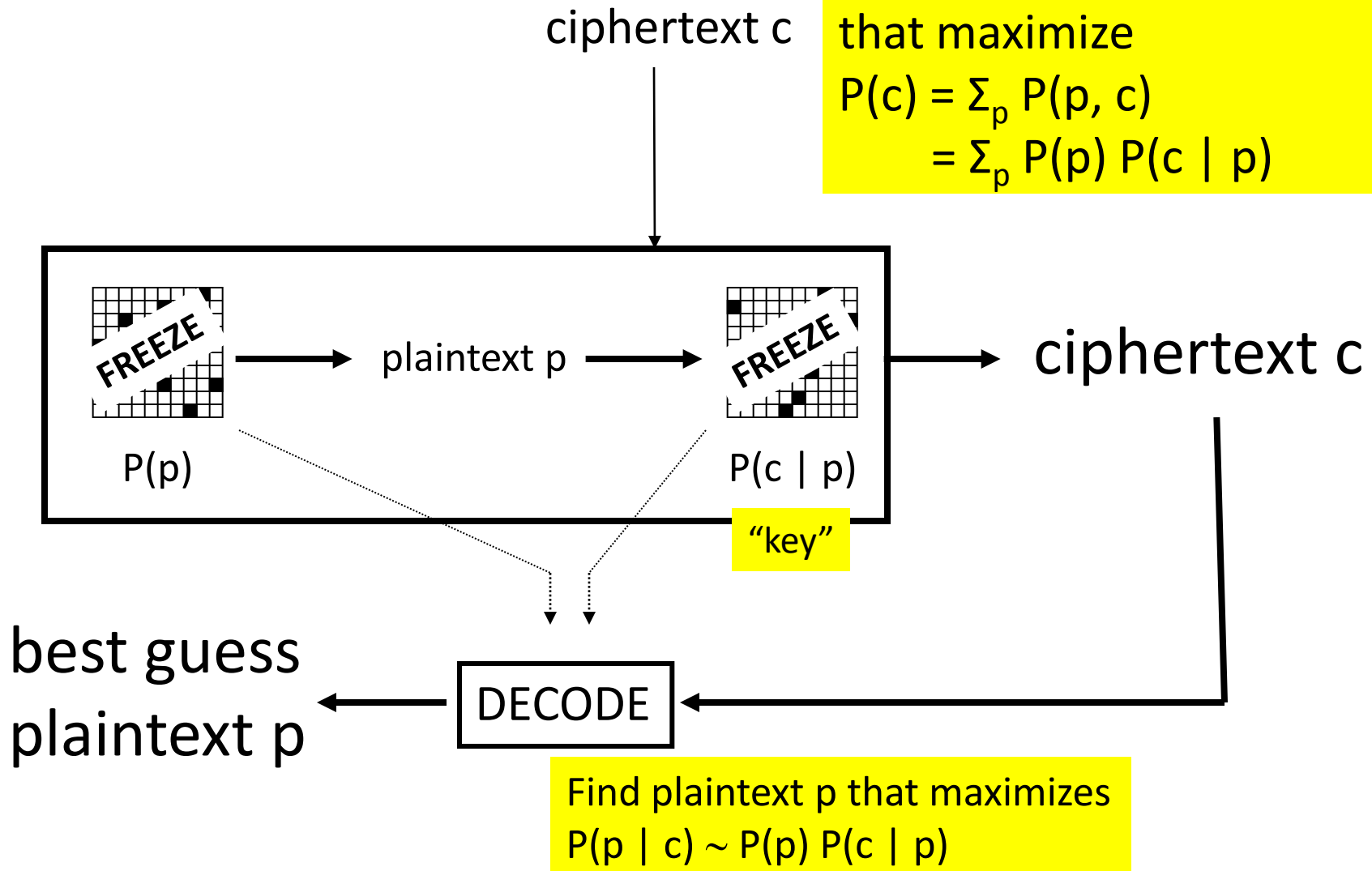
$$\begin{aligned} P(c) &= \sum_p P(p, c) \\ &= \sum_p P(p) P(c \mid p) \end{aligned}$$



Letter Substitution Cipher

Find substitution-table values that maximize

$$P(c) = \sum_p P(p, c) \\ = \sum_p P(p) P(c | p)$$



Letter Substitution Cipher

plaintext samples,
unrelated to ciphertext

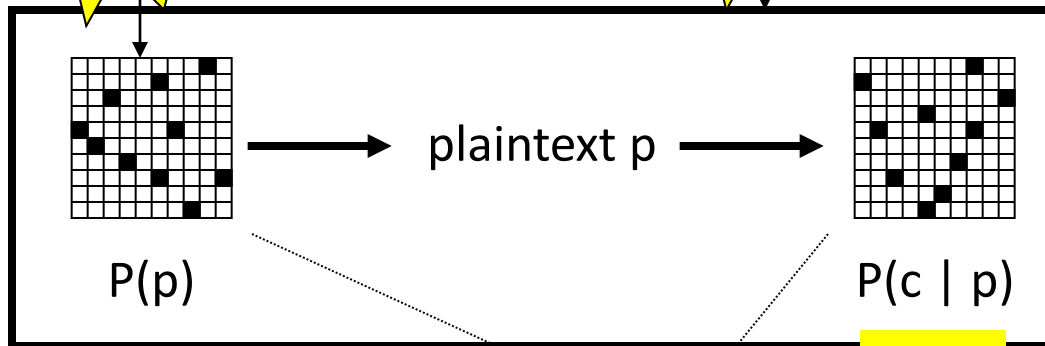
LM

ciphertext c

EM

Find substitution-table values
that maximize

$$P(c) = \sum_p P(p, c) \\ = \sum_p P(p) P(c | p)$$



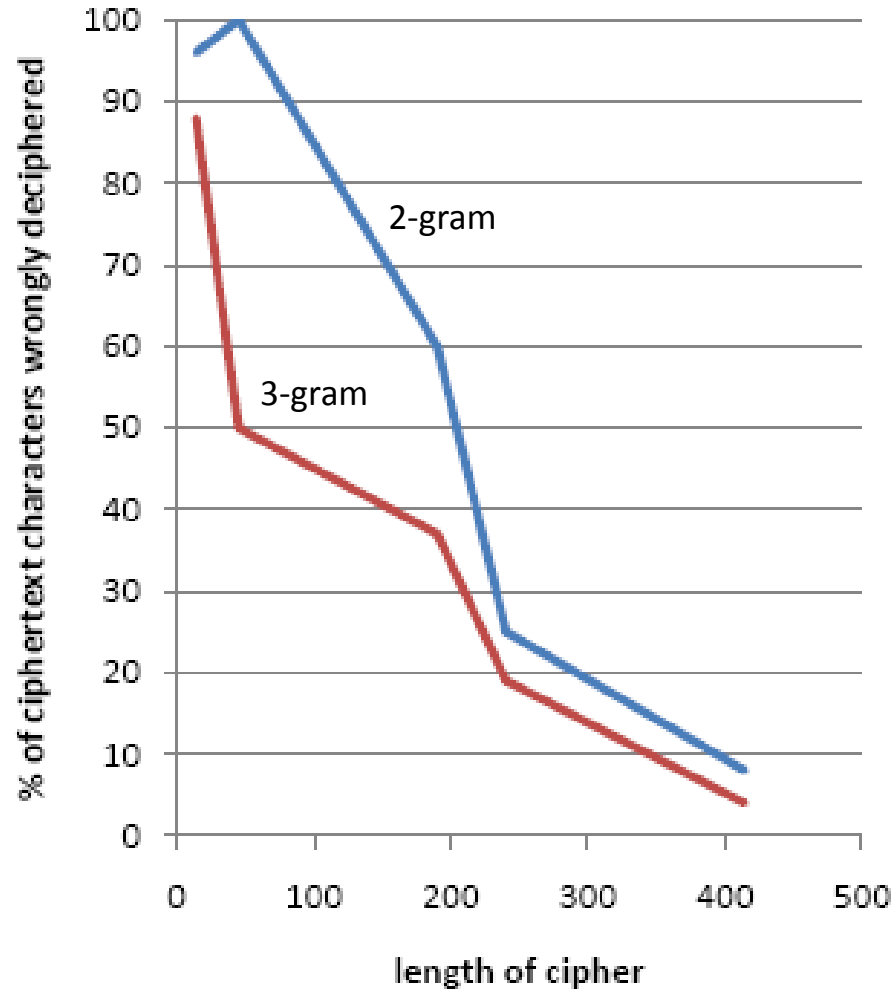
"key"

best guess
plaintext p

Viterbi

Find plaintext p that maximizes
 $P(p | c) \sim P(p) P(c | p)$

Decipherment Accuracy vs. Cipher Length



Letter Substitution Cipher

plaintext samples,
unrelated to ciphertext

LM

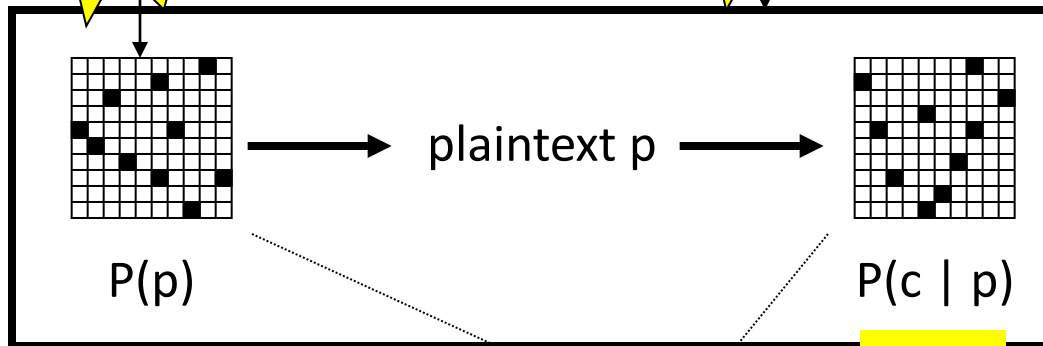
ciphertext c

EM

Find substitution-table values
that maximize

$$P(c) = \sum_p P(p, c) \\ = \sum_p P(p)^{0.5} P(c | p)$$

[Ravi & Knight 09b]



ciphertext c

best guess
plaintext p

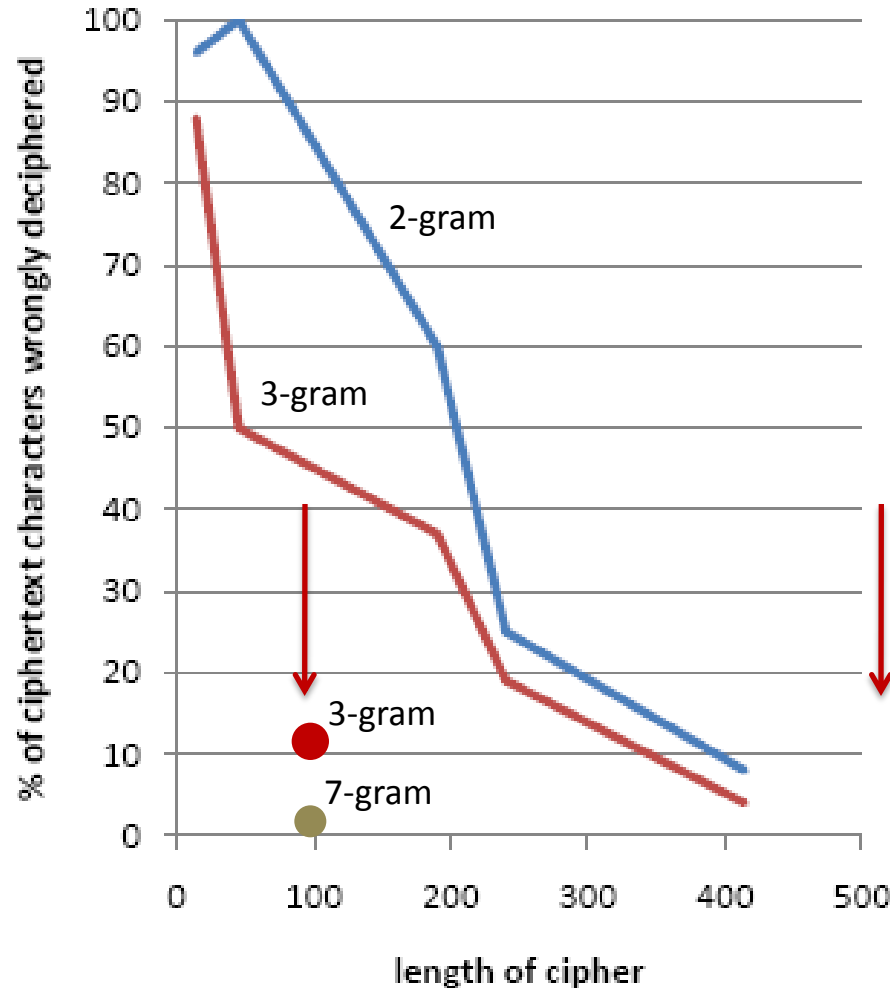
Viterbi

Find plaintext p that maximizes

$$P(p | c) \sim P(p) P(c | p)^3$$

[Knight/Yamada 99]

Reducing LM Weight During EM



Set EM to maximize

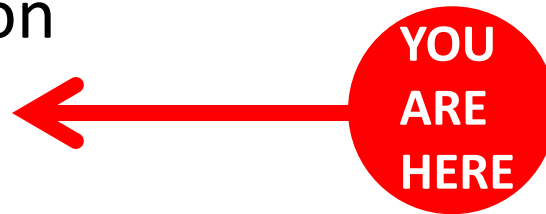
$$P(c) \approx \sum_p P(p)^{0.5} P(c | p)$$

instead of

$$P(c) \approx \sum_p P(p) P(c | p)$$

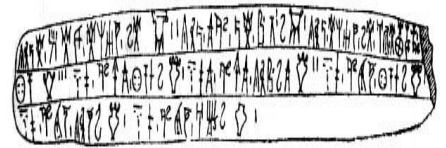
Plan for This Talk

- Break a series of codes
 - Simple letter substitution
 - Phonetic substitution
 - archaeology
 - transliteration
 - Word substitution
 - Foreign language as cipher
- Bonus
 - Two historical ciphers
 - Final thought on translation and cryptography



Phonetic Decipherment

ciphertext



Phonetic Decipherment

ciphertext

**primera parte
del ingenioso
hidalgo don ...**

Phonetic Decipherment

“When I look at these squiggles, I say to myself, this is **really a sequence of Spanish phonemes**, but it has been encoded in some strange symbols...”



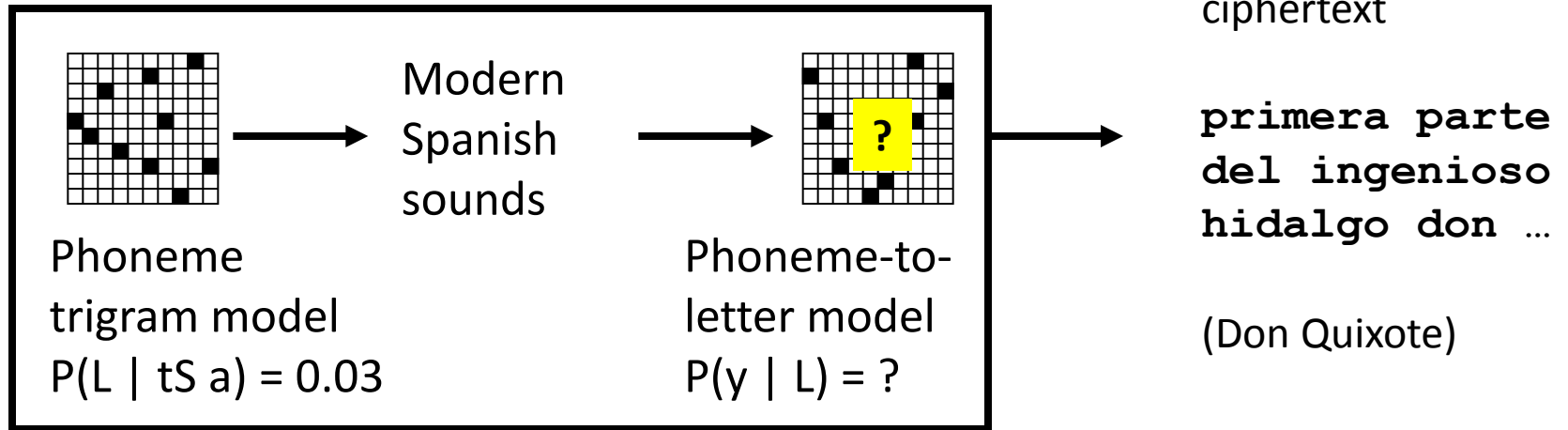
OUR HERO

ciphertext

**primera parte
del ingenioso
hidalgo don ...**

(Don Quixote)

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:

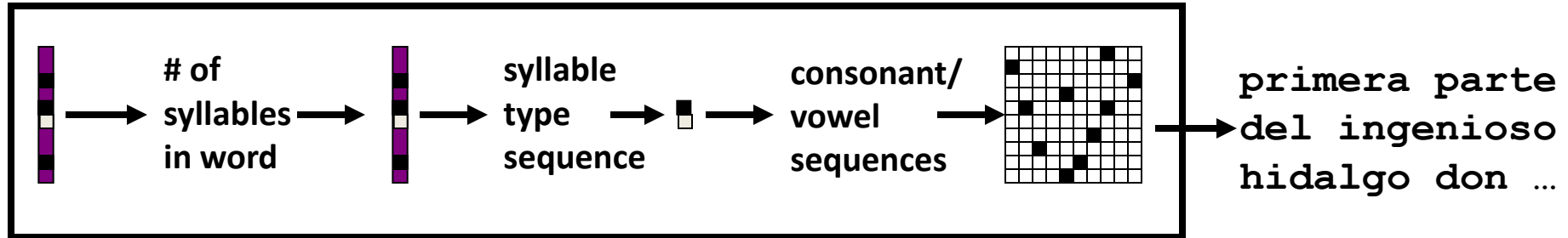
ñ, á, é, í, ó, ú,
 a, b, c, d, e, f, g,
 h, i, j, k, l, m, n,
 o, p, q, r, s, t, u
 v, w, x, y, z

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 $P(c \mid p)$ 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

Unknown Source Language



$P(1) = ?$
 $P(2) = ?$
etc.

$P(CV) = ?$
 $P(V) = ?$
 $P(CVC) = ?$
+ 7 others

$P(V | V) = ?$
 $P(VV | V) = ?$

$P(a | V) = ?$
 $P(a | C) = ?$
etc.

Input: **primera** **parte** **del** **ingenioso** ...
Output: **NSV.NV.NV** **NVS.NV** **NVS** **VS.NV.SV.V.NV** ...

S = sonorous consonant phoneme

N = non-sonorous consonant phoneme

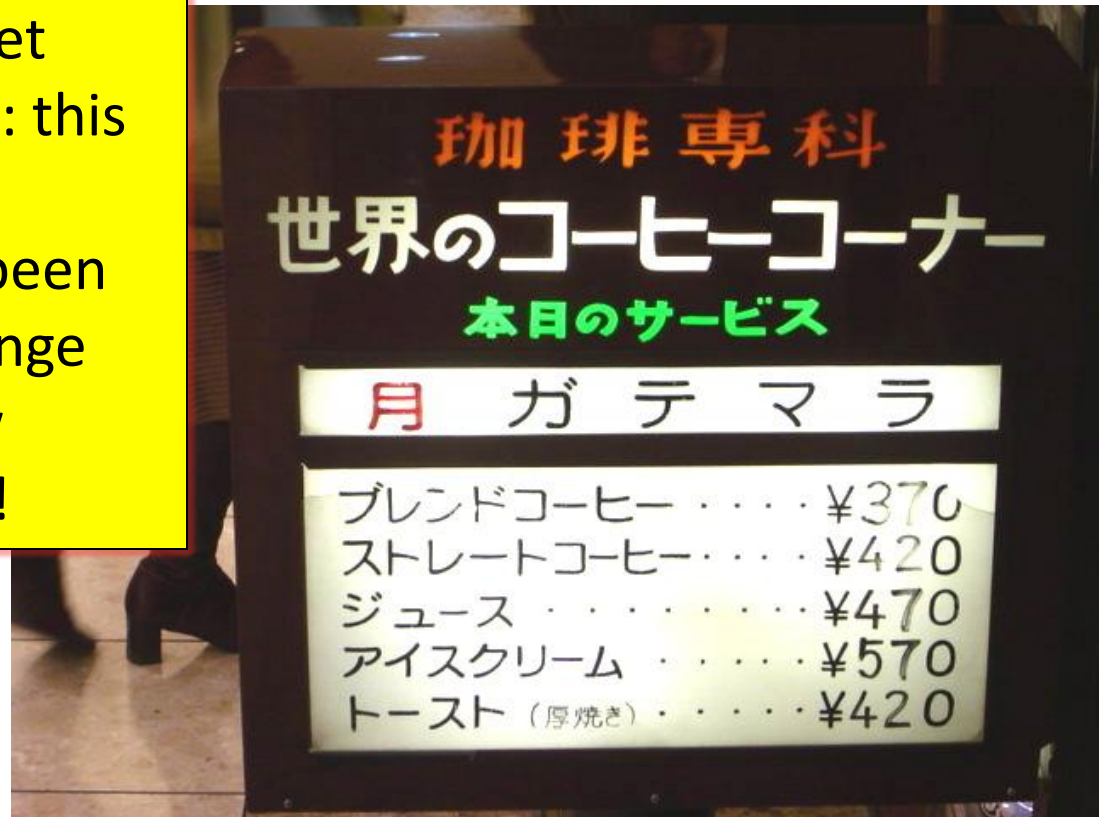
V = vowel phoneme

Phoneme Substitution Ciphers

When I look at street signs in Tokyo, I say: this is **really written in English**, but it has been coded in some strange symbols. I will now proceed to decode!



OUR HERO



Parallel data:

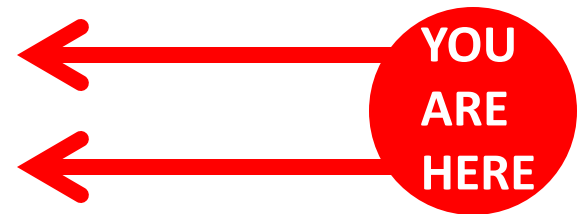
[Knight & Graehl 97]

Non-parallel data:

[Ravi & Knight 09a]

Plan for This Talk

- Break a series of codes
 - Simple letter substitution
 - Phonetic substitution
 - archaeology
 - transliteration
 - Word substitution
 - Foreign language as cipher
- Bonus
 - Two historical ciphers
 - Final thought on translation and cryptography

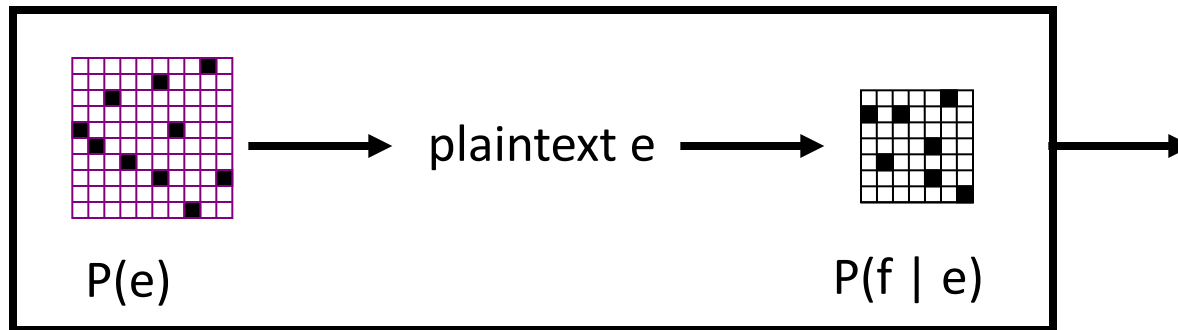


Foreign Language as a Cipher

“When I look at **this giant corpus of Arabic**, I say to myself, this is really English, but it has been encoded in some strange symbols!!! Let's decode!!!”



OUR
HERO



رفض رئيس السلطة الفلسطينية محمود عباس مجددا تصريحات وزير الخارجية الإسرائيلي سيلفان شالوم التي قال فيها إنه يتعين 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 يتعين 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 الجانب للانسحاب الإسرائيلي من غزة، وأكد أن إسرائيل تريد مغادرة هذه الأراضي 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 الجانب

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

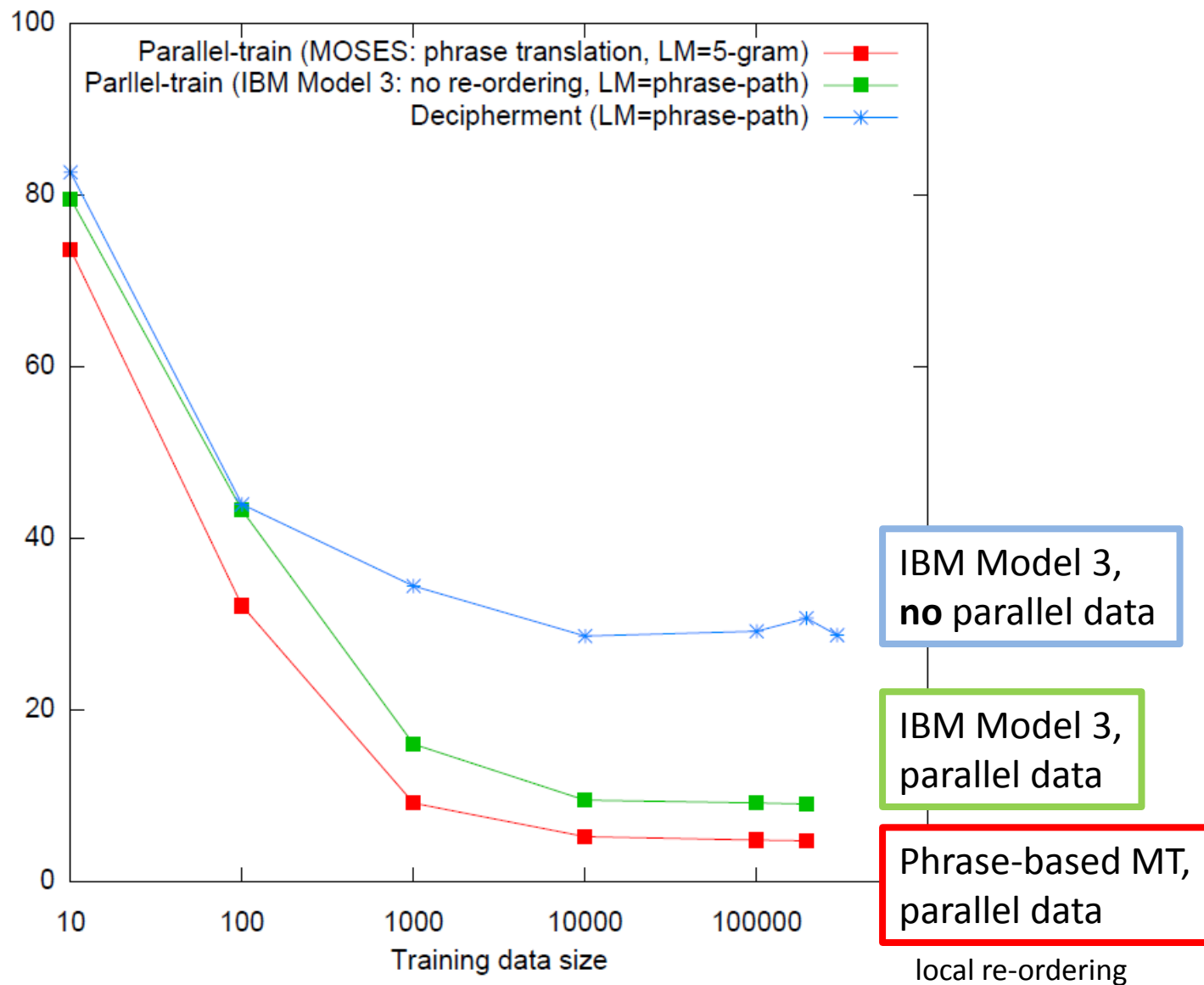
229 fy Hzyr!n !lm!Dy
207 fy Hzyr!n
75 fy Hzyr!n !lmqbl
61 fy Hzyr!n 1993
31 fy Hzyr!n 1992
27 !lr!b@ mn Hzyr!n
27 fy Hzyr!n 1967
19 fy 30 Hzyr!n !lm!Dy
18 fy n=h!y& Hzyr!n !lm!Dy
18 fy Hzyr!n 1991
17 mn Hzyr!n
17 mndh Hzyr!n !lm!Dy
17 4 Hzyr!n

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

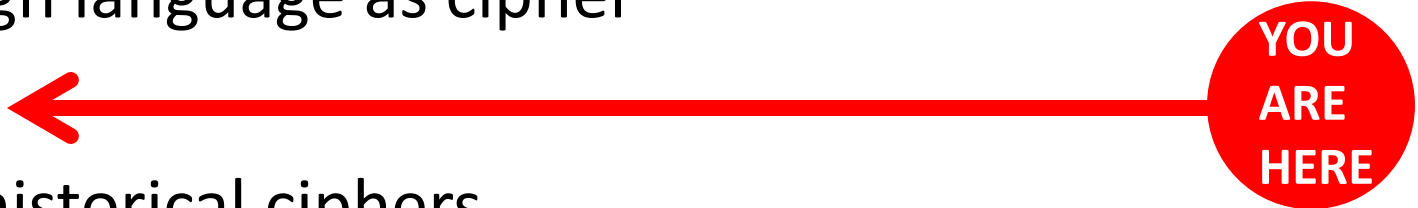
Deciphering Spanish Time Expressions

MT quality
on test set

(Edit distance,
lower is better)



Plan for This Talk

- Break a series of codes
 - Simple letter substitution
 - Phonetic substitution
 - archaeology
 - transliteration
 - Word substitution
 - Foreign language as cipher
- Bonus 
 - Two historical ciphers
 - Final thought on translation and cryptography

Copiale Cipher

=mzfzhi|ny+cz||nēmlēq||iz+rbgymk|sōyλ+ab+qanpāz
 h|icjyānu+qalp+rgdpyx|dāyλagdrf+amz+zcūh|rprir|dēyλ
 e. h=anūir|dōqz||agōivn|t+bz|ōs|sēq||m=rōmz||būid:q|p|o
 cō||ar|t|ndg+pucl|p|k|λ|ēλ|u|o|d||:λ|b|λ|n|p|u|b|z||:ō|λ|b|y|h|m|g|r|p
 m|pōr|u|o|d||n:sēq|p|u|+ū|j|d|+n|h|c|u|g=r|z|g|m|Δ|λ|u|lāy|g||w+q|x
 j||m=m|p|ac|t|h|z|c|r|x|i=g|z|n|cō||λ|b|t|h|+v|z|y||ē|b|g|n|r|m|z|i|g|z|+h|ā|l
 h=|w|z|ē|w|λ|b|c|m|d|u|d|h|k|n|p|b|h|r|=+p|m|r|+h|r|u|i|o|l|z|n||n||j
 g|n|+v|u|z||h|z|i|m|+r|ē|+Δ|r|ē|c|p|c:|p|r|n|q|r|m|p|ā|o|l|p|ū|+v|z|p|o|f|j|u
 p|u|λ|h|s|=l|ō|r|ā|j|r|ē|r|u|b|f|h|u|p|c|ē|r|p|λ|e|z|n|n:f|c|y|ē||λ|w
 z|j|ā|λ|b|z|y|ā|ū|m|h|j|r|ā|l|o|=g|z|g|+h|g|t|x|ū|h|λ|h|r||b|h|t|+j|λ|ō|x|s|+|=p|z
 a|=+g|z|q|+m|o|d||x|m|p|h|ē|y|λ|v|z|ā|b|c|j|u|p|ā||f|q|z|u|o|p|t|w|ū|r|ō|c|y|f|u
 d|x|ō||λ|az|ū|v|p|ō|d|=λ|r|Δ|λ|u|v|x|j|Δ||λ|v|h|g|r|z||+b|Δ|r|λ|n|p|ā|z
 m|r|=o|f|ō|p|r|m|Δ|m|z|q|ā|h|Δ|n|f|u|p|+|v|λ|ō|Δ|+b|f|=h|c|u|p|t|r|g|z|h|ē
 h=|r|r|h|+ā|b|ō|r|r:|r|λ|ō|λ|u|+|q|h|ē|z|b|ō|Δ||f|ū|h|ē|z|o|z|h|r:|l|p|ō|g|g|=u|w
 p|j|z|ā|r|ga=g|p|v|ū|z|z|ē|q|r|z|h|c:ū|r|g|h|r|m|ō|f|u|z|u|w|r|z|q|p|ū|r|f|=g|z
 g|p|ō|j|u|v|~~h|c|+n|z|x|h|θ|z|h|o|u|v|p|r|+|λ|o|Δ|d|+h|m|c|f|m|j|f|z|n~~
 o|n|ū|m:ē|λ|o|s|=l|j|y|r|d|u|e|m|r|r:|d|h|ā|z|g|q|r|m|r|c:||v|d|Δ|λ|ō|Δ|c:q

$$3\hat{a}A$$

3410 murem[m]b[+]c[+]p[+]k[+]x[+]z[+]d[+]b[+]c[+]x[+]m[+]z[+]h[+]o[n] Δ[+]d[+]d[+]e[+]f[+]c[+]a[+]u[+]a[+]d[+]y[+]e[+]c[+]a[+]i[+]p[+]a[+]n[+]h[+]z[+]p[+]r[+]o[+]y[+]m[+]p[+]b[+]p[+]h[+]e[+]s[+]a[+]n[+]x[+]h[+]i[+]c[+]u[+]b[.]

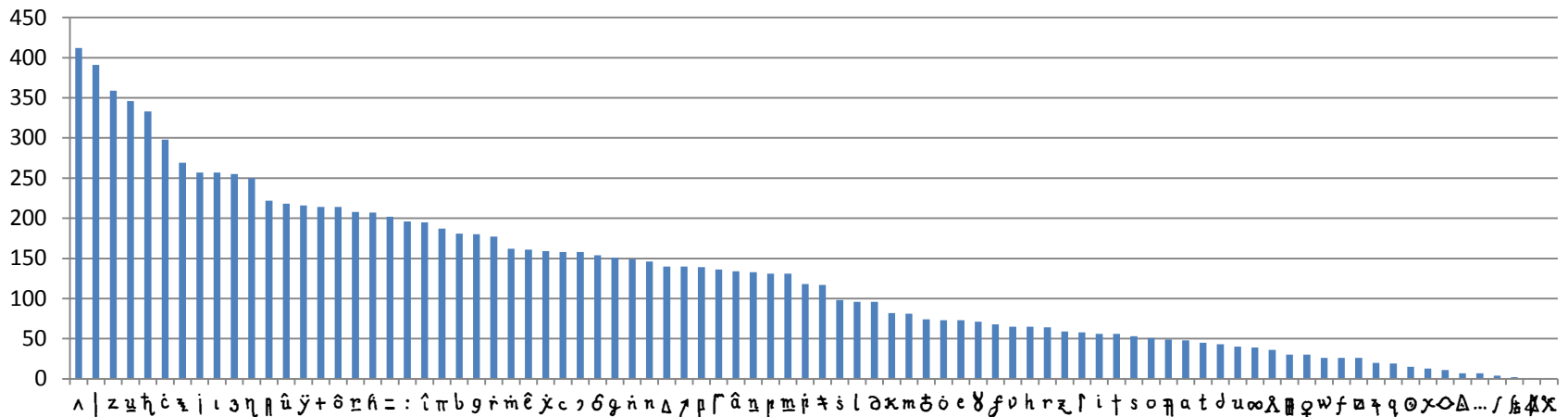
Ирѣво+ѡхѣ/ѣ:ѡв.

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

$\text{Ar}^{\text{e}}\omega + b \cdot \text{In}.$

Σαρκοειρήνην ἄρ. Ἀναΰχινος ἔργ. = 39 h 17

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:

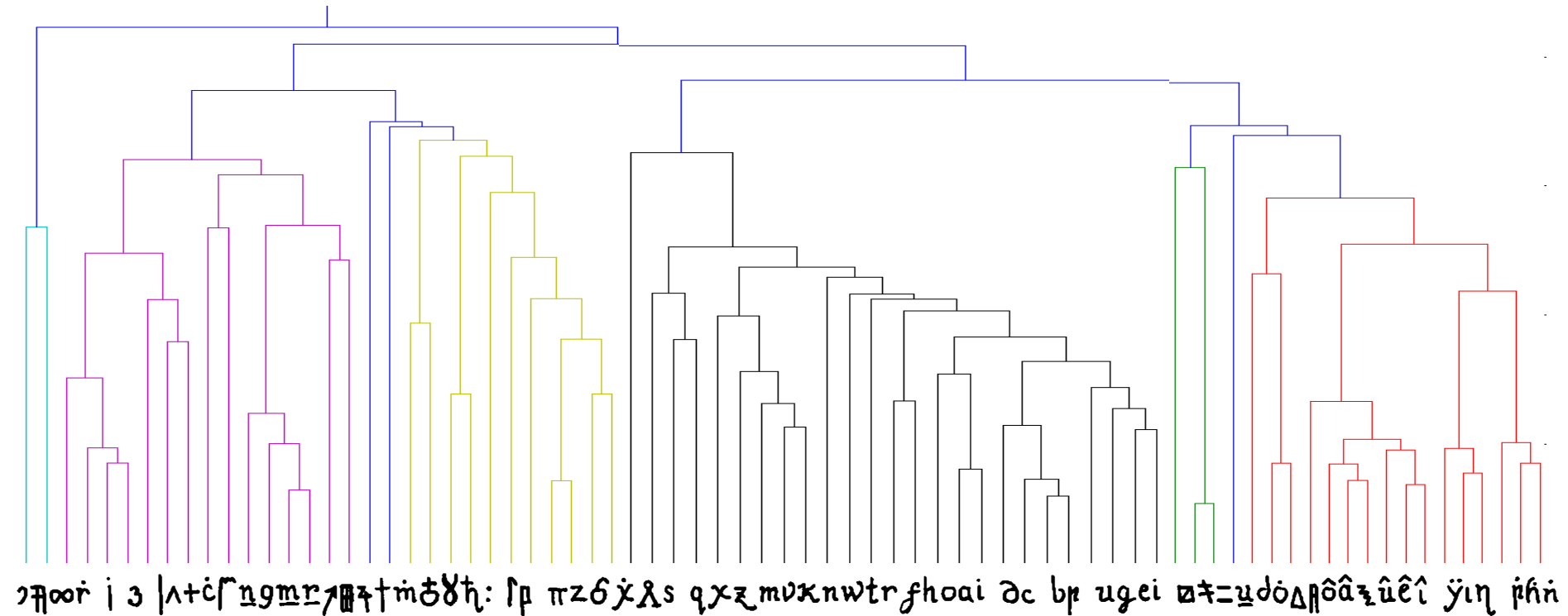
â, ê, î, ô, û followed by **3** and **j**

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

Clustering of Cipher Letters

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

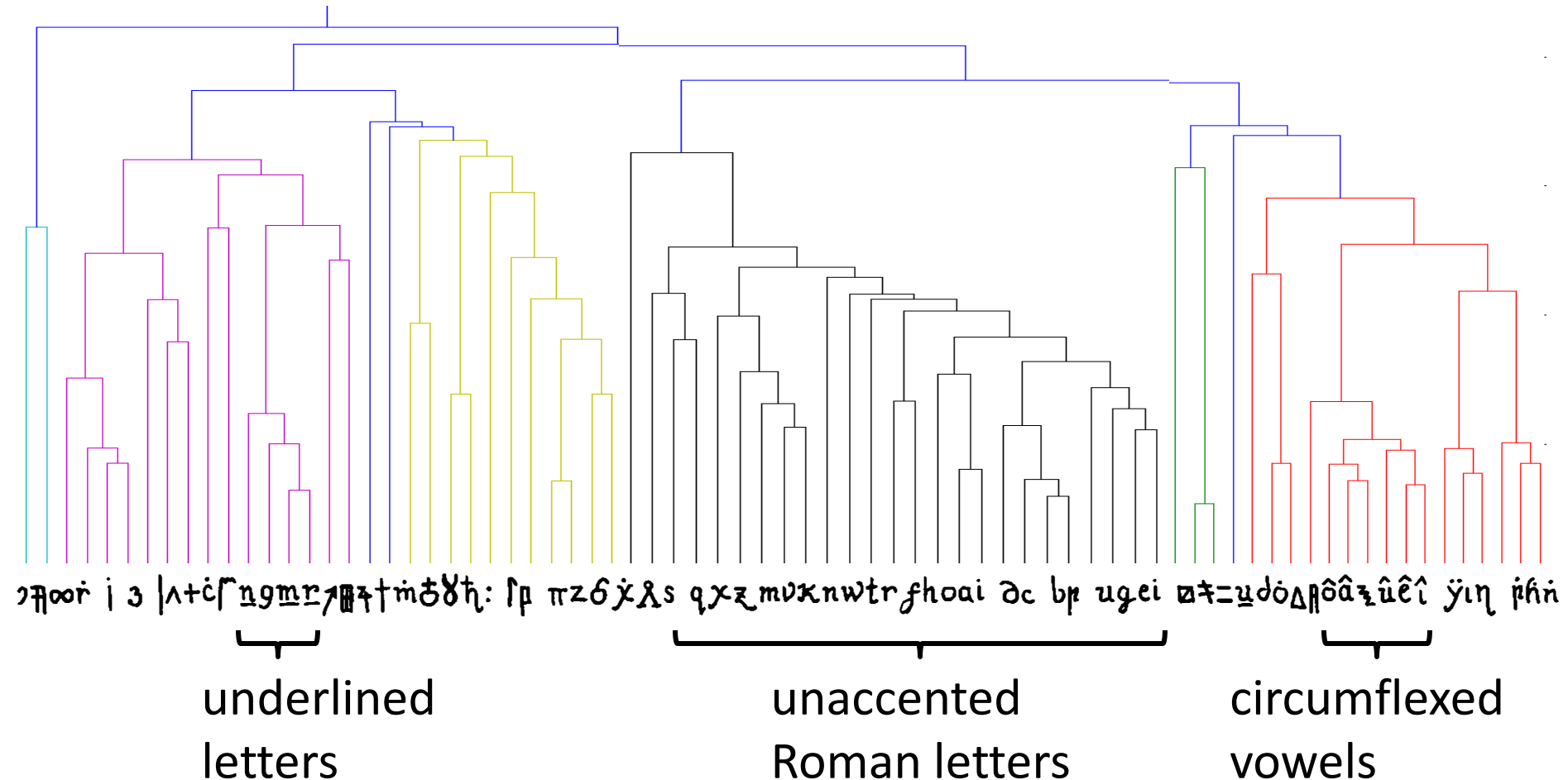
Scipy software



Clustering of Cipher Letters

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

Scipy software



thanks Jon Graehl

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 η 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 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 ...

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 = | 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 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 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 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 | 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 ...

D, 80, **FAIL** nst
... languages.

Second Decipherment Approach

Homophonic cipher,
e.g.:

A = ʒ j l y r

B = û

C = ô ñ

D = ʈ

E = ʁ ʃ Δ * f î ɜ ɔ

F = ʀ

G = ȳ

etc.



κ m û r : ʀ z i ô ʃ | ȳ ɔ ʔ ê j ʔ ɜ i ʌ n ʈ â ɜ b Δ g z =
i ʀ l z u ʀ ɔ ç ʌ â r ɔ κ ʌ ʔ ʈ r ʔ ʈ ʌ ɜ i n ʀ î | Δ r ô ʀ ʌ ʌ
= g z w ʈ ȳ ê c Δ r ô Δ + b z ɣ r i ʁ ȳ j î r z u ʃ ʌ ɜ
ʈ * j ʀ ɔ ʀ = ʃ | û ʌ s * m ô m | ɜ ɣ g â | κ ʔ = ʌ ʔ | l ʁ ô
ø ʃ : r î ʌ b i ʃ u m ȳ j z v z â j ʁ ɔ ʀ m ʈ i z ʁ ʌ c ô ô g g
z û + ʀ * ʈ ʈ ʁ ê ɜ g h ʌ ʔ ʈ i ʔ ʈ ʌ î ɜ t n * | r ô ȳ m â
+ h ʁ r z ô ɜ ñ b s ɣ + : ɜ r κ ʀ ʀ ô ʀ ʔ Δ c û ʌ g = ñ z κ
ʀ ɜ ∞ n z ʀ ɜ ʃ ʔ ʈ r ʈ z ê ȳ ñ g = r ʈ g ô Δ ɜ n z ʀ | κ
ʈ ɣ j i ʁ ȳ r î g ʈ u ʌ b g ʌ i ʈ b ô û ʃ ʈ ʔ ɜ ô ʌ e z ɣ ô
| û r c ʃ ɜ ɔ ô ʌ b ɣ ʔ m ɔ

Homophonic Cipher

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

FAIL

But, slight numerical preference for
German

Cipher Characteristics

digraphs:

ʁ ʰ 99

č : 66

ʰ ^ 49

: ȳ 48

z ʀ 44

trigraphs:

ʁ ʰ ^ 47

č : ȳ 23

ŋ ʁ ʰ 22

ȳ ʁ ʰ 18

ʁ 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

Key Observation #1

In Copiale, ʔ almost always followed by ᵀᵇ

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

So guess: ʔ = C, ᵀᵇ = H

One Thing Leads to Another

$$\text{ʝ}^{\text{t}}\text{h} = \text{CH} \quad \rightarrow \quad \text{ʝ}^{\text{t}}\text{h}\text{ʌ} = \text{CHT} \quad \rightarrow \quad \text{ʌ} = \text{T} ?$$

Each step is guesswork.

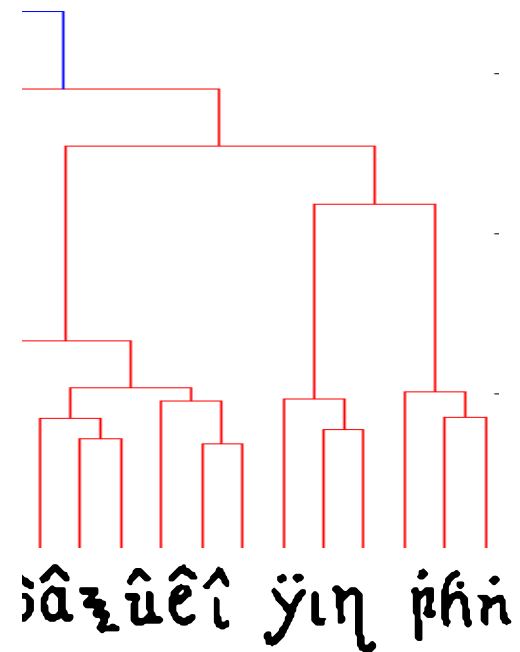
Must be willing to retract.

Weird task, not knowing German.

No longer care what the book says.

Cluster diagram crucial:

$$\text{ȳ} = \text{ɪ} \quad \rightarrow \quad \text{ɪ} = \text{ɪ} , \text{ɳ} = \text{ɪ}$$



Spring Break 2011

c aeiou fpy dlmrztbvw hkngs j

german ^{aus} ub

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

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 ʊ

Δ o p h

3 m c r + = #

need: f g y l m z

(rare on
 german)
 u f
 a f
 r f
 p f

a o i
 v w
 k g j

k e
 g e
 j e

h g

german {u r i} u {c b d a p i g l}

δ x

z u

a u f

u n

u r

h u

3 i r

l w i g o p r d l

o

o

o

o

3 i r

r m

* der ✓ eht 2 h ^
 und ✓ e i : u
 ein ✓ s l h i l
 ung ✓ ich 2 h
 eht ✓ ich 2 h
 ich ✓ che 2 h
 sch ✓ t ^ m Δ
 che ✓ t i c l ^
 ech ✓ e t t g ^
 * die ✓ s a h =
 rec ds 2 h
 ine e l t i ^
 gen t c l n
 est ✓ t e l i u
 ver t i l i ^
 hen der 2 h
 lic der 2 h
 ten ew u a i
 rei sch h 2 h
 nke st m k l
 auf ich l 2 h
 ede ein h 2 h
 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 u t m ^
 hei und = n
 nsc z m Δ i
 ens i h c
 men tot ^ h
 hat o h 2
 ere i c
 * das a k t m
 rde ste h a u
 nte s h c
 nge sch h 2 h
 lte che c i
 ore che 2 h u
 ede t l a c

Spring Break 2011

German letters

Cipher letters, in groups

Grid

vowels: u ʔ ɛ ɪ ɔ ɯ

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

need: f o y l m z

(repeated)

| | |
|----|----|
| uf | sp |
| af | pr |
| rf | pe |
| ff | |

aoi
vw k g

K_e Ge
 K_a Gr

$$Z[u] \quad \{sv_i\} \cup \{nr, dl\}$$

| | | | |
|---|---|---|---|
| a | u | f | 0 |
|---|---|---|---|

| | | |
|---|-----------|---|
| u | v | 0 |
| u | \bar{v} | 0 |

fu

218

211 315

$\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$

| | | | | | |
|-----|---|-----|---|---|---|
| der | ✓ | cht | 2 | h | u |
| und | ✓ | e | 1 | : | u |
| ein | ✓ | s | 1 | h | i |
| una | | ich | g | 2 | h |
| cht | ✓ | ich | n | 2 | h |
| ich | ✓ | che | 7 | h | u |
| sch | ✓ | t | 1 | m | a |
| che | ✓ | i | 1 | l | a |
| ech | | it | + | g | 1 |
| die | ✓ | s | + | h | 1 |
| rec | | ds | 7 | h | 1 |
| ne | | el | + | l | 1 |
| gen | | t | 1 | l | 1 |
| ent | ✓ | t | 1 | l | 1 |
| ver | | ti | 1 | l | 1 |
| hen | | der | = | h | 3 |
| lic | | der | = | h | 3 |
| ten | | ew | u | a | i |
| rei | | | h | a | i |
| auf | | sch | h | a | i |
| ede | | st | h | a | i |
| and | ✓ | ich | 1 | 7 | h |
| don | | ein | h | g | 1 |
| run | | t | 2 | h | 1 |
| ter | | ie | 2 | h | 1 |
| ere | | die | 2 | h | 1 |
| sei | | e | 2 | h | 1 |
| hte | | e | m | 1 | 1 |
| hei | | u | m | 1 | 1 |
| ns | | und | = | h | 1 |
| ens | | z | m | 1 | 1 |
| men | | i | h | 1 | 1 |
| hat | | tet | 1 | h | 1 |
| ere | | h | 1 | h | 1 |
| das | | g | 1 | h | 1 |
| nde | | ste | h | 1 | 1 |
| nte | | s | h | 1 | 1 |
| ge | | sch | h | 1 | 1 |
| te | | che | 2 | h | 1 |
| ere | | | 2 | h | 1 |
| ode | | t | 1 | h | 1 |

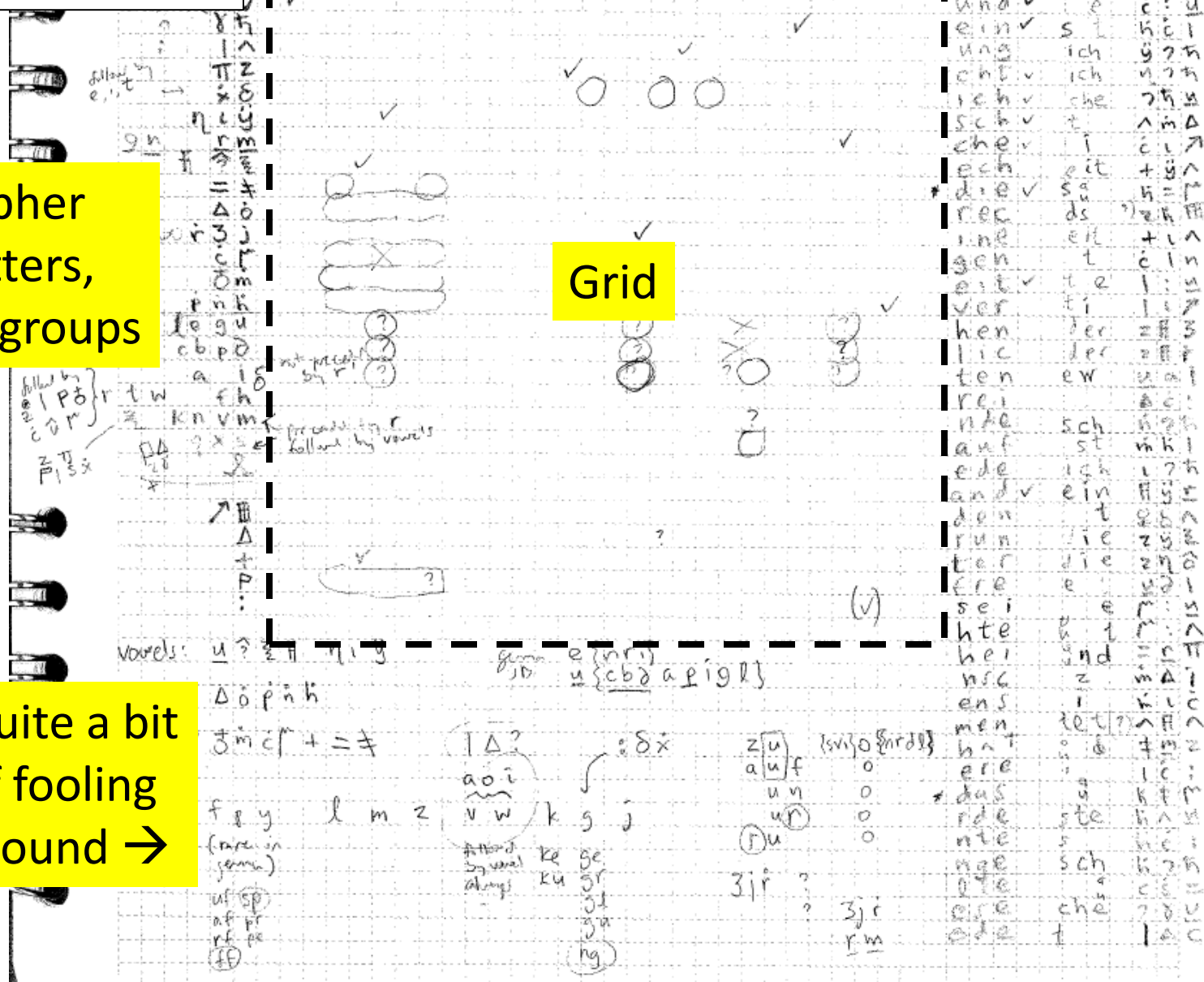
Spring Break
2011

German letters

Cipher
letters,
in groups

Grid

Quite a bit
of fooling
around →



Spring Break 2011

German letters

German
trigraphs

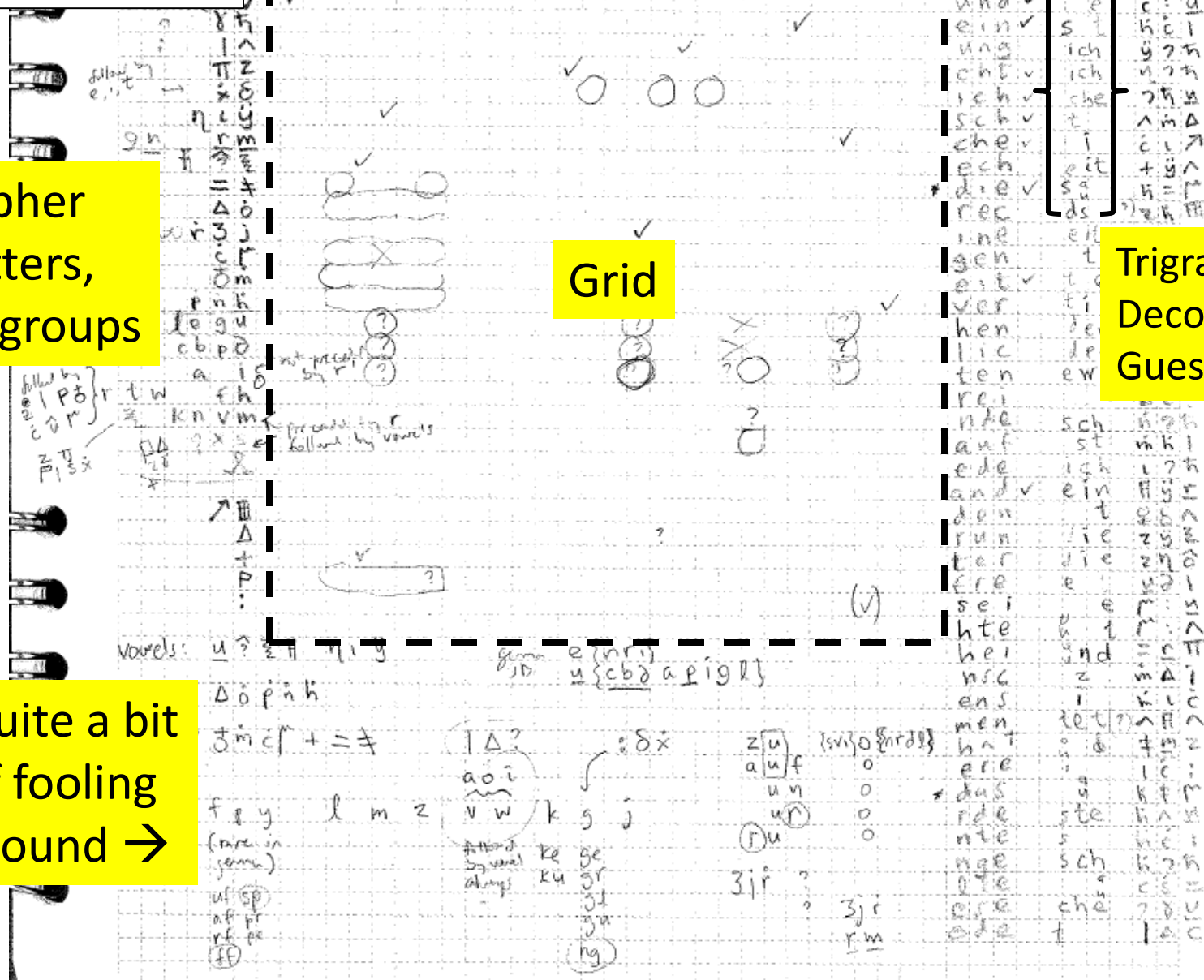
Cipher
trigraphs

Cipher
letters,
in groups

Grid

Trigraph
Decoding
Guesses

Quite a bit
of fooling
around →



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

Actually, those are space bars

Copiale Decipherment

lit:mz||bl

υχζ|ιλςκρξγ|ωη

πδ|ρτδΔρεζcâ=γλûb ◊ ur ⊙ z

δρτξι+δ|ιρλτλ̂ηĉf.

cû|fēztr̂p†gηλ:κ

δχûτλ̂η+ξ̂r̂κ̂m̂λ̂ι̂z:γ̂λ̂δ̂ôiq̂ẑι̂χ̂â|ēĉ:υ̂δ.

fûr̂f̂ôĵκ̂λ̂ι̂λ̂=̂ĉz.

m̂ôr̂â+Δĝγ̂υ̂x̂ẑδ̂m̂n̂κ̂f̂n̂ĥτ̂+̂îf̂.

κ̂m̂û̂r̂:π̂ẑι̂δ̂f̂|γ̂ôτ̂ē̂ĵτ̂ξ̂ι̂λ̂n̂π̂â̂ẑb̂Δ̂ĝẑ=̂ĵλ̂ẑυ̂ρ̂ĉλ̂â̂r̂ĝκ̂λ̂τ̂

π̂r̂τ̂π̂λ̂ξ̂ι̂ρ̂ĉ|Δ̂r̂δ̂ρ̂λ̂â=ĝẑŵπ̂γ̂ēĉΔ̂r̂δ̂Δ̂+b̂ẑη̂r̂ι̂χ̂γ̂ĵĉr̂ẑυ̂f̂λ̂ẑ

π̂κ̂ĵr̂δ̂r̂=̂f̂|û̂λ̂ς̂κ̂m̂δ̂m̂|ξ̂η̂ĝâ|κ̂τ̂=̂λ̂τ̂|l̂χ̂δ̂ω̂f̂:π̂λ̂b̂ι̂f̂ûm̂γ̂ĵẑυ̂

ẑâ̂ĵx̂r̂m̂π̂ι̂ẑĥλ̂ĉδ̂ôĝĝẑû̂+̂r̂κ̂q̂n̂χ̂ē̂ẑĝĥλ̂τ̂π̂ĵτ̂π̂λ̂îẑt̂n̂κ̂|r̂δ̂γ̂

m̂â̂+ĥĥr̂ẑδ̂ẑn̂b̂ŝη̂+̂:ξ̂r̂κ̂r̂δ̂ρ̂τ̂δ̂ĉû̂λ̂ĝ=̂n̂ẑκ̂ρ̂ξ̂ôon̂ẑρ̂ẑf̂τ̂ĥr̂π̂

ẑē̂γ̂n̂ĝ=̂r̂π̂ĝδ̂Δ̂ẑn̂ẑρ̂|κ̂π̂η̂ĵι̂χ̂γ̂r̂îĝπ̂υ̂λ̂b̂ĝλ̂ι̂f̂b̂δ̂û̂f̂π̂ĵẑδ̂λ̂ĉ

ẑη̂δ̂|û̂r̂ĉf̂ẑδ̂δ̂λ̂b̂η̂τ̂m̂δ̂:

n̂îr̂f̂ĉîr̂êôp̂ĥδ̂r̂δ̂p̂ξ̂û̂τ̂ĥẑâ̂κ̂ ◊ ĝẑ=̂δ̂m̂ē̂ĵẑυ̂l̂î

ĝẑm̂û̂ôôλ̂δ̂n̂|ôπ̂ē̂ĝυ̂δ̂δ̂ξ̂r̂Δ̂ĵẑĝκ̂r̂χ̂υ̂ĝπ̂û̂ẑκ̂ ⊙ n̂|η̂τ̂b̂=̂n̂

λ̂ē̂r̂m̂δ̂ẑf̂:υ̂â=̂r̂ẑl̂Δ̂τ̂r̂ρ̂δ̂m̂η̂â̂ẑδ̂ĵ|δ̂îr̂f̂δ̂λ̂δ̂γ̂λ̂ôẑι̂ē̂ĝĉû̂τ̂

r̂ŝē̂η̂λ̂

ĉr̂=̂f̂|â̂τ̂υ̂b̂m̂Δ̂ĉ:ξ̂δ̂.

ĥẑĵη̂λ̂:δ̂ĝ|êẑη̂â̂ẑ Δ̂n̂π̂â̂ẑ ⊙ p̂ς̂κ̂r̂δ̂îẑt̂m̂â̂γ̂δ̂υ̂l̂î=̂n̂π̂

ẑr̂ŝ=̂n̂ĥκ̂f̂r̂ẑr̂|n̂δ̂ξ̂ĵp̂ûĝπ̂ĉγ̂ôτ̂ĥf̂δ̂b̂|r̂κ̂n̂ĝτ̂ξ̂γ̂t̂ι̂δ̂ĉŝ=̂b̂+̂n̂γ̂υ̂δ̂

χ̂δ̂|r̂:û̂λ̂ô|ôon̂.

l̂π̂δ̂ẑf̂ôĥĝẑγ̂p̂r̂λ̂n̂m̂λ̂m̂Δ̂r̂λ̂ôêκ̂n̂r̂.

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 dirigirenden 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

dirigirenden λ tisch geführt dieser fragt ihn:

erstlich ob er begehre ◊ zu werden

zweytens denen verordnungen der ⊙ sich

unterwerffen und ohne widerspenstigkeit die lehrzeit

ausstehen wolle.

drittens die Δ der ⊙ gu verschweigen und dazu

auf das verbindlichste sich anheischig zu machen

gesinnet sey.

der candidat antwortet ja.

Copiale Decipherment

lit:mz||bl

υχζ|ιλςκρξ|wn

πoιρhΔxē3cā=γλûb ◊ ur ⊙ x

δρhξι+οιnrλhιηc f.

cû|fē3tîp†gηλ:κ

θχûhēη+ξrκmλi3:γλδoιqziîχâ|ēc:υθ.

fûrδoικλiλ=cz.

mγrâ+Δgÿuxzδ3mñκfñhî+îf.

κmûr:μziδf|ÿo hēj hξiλnπâ3b Δgz=|zlzμpôcλârgκλh

πr hξiλnπrî|δrδrλa=gzwπÿēc ΔrδΔ+bzηrιχÿjîrzu fλx
πκiγδr=|ûλςκmδm|ξηgâ|κh=λh|j|χδωf: rîλbιf umÿjzυ
zâjxγr mπiz hλcδôggzû+μκπñχē3ghλhπihλi3tñκ|rôÿ
mâ+hñr zδ3nbsη+: ξrκrδrδhδcûλg=ñzκpξoonzλ3 f hñπ
xēÿñg=rπgδΔ3n3r|κπηjιχÿrîgπu λbgλι†bδûf h h3δλc
zηδ|ûr c f 3ôδλbη h mδ:

ηîr fci γεôphôrδpξδûh3âκ ◊ g3s=δmējzuli

g3mûoolδn|oπēgυδθξrδjz9κrξu gπû3κ ⊙ n|ηr b=ñ
λêr mδ3f: uα=rzλΔ h rπδmηâzδi|δîr f iδ8ôÿλoz iēg cûh
rîsēηλ
cî=|fâ h u b mδc: xδ.

hziηλ:ôg|ezηâx λ nπâ3 ⊙ pîs κrδi3†mâÿδu li=ñπ
zîs=ñhκf r zî|nδξiμgπcÿo h fδb|jκñg h ξÿ†iδcs=b+ñγuδ
χδi r: ûλv|îoon.

lπδ3 fγhgzÿπrλlñmλmδrλrλocξñr.

First lawbook

of the ◊ e ⊙

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 λ, 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 λ, who asks him:

First, if he desires to become ◊.

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

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

The candidate answers yes.

Copiale Decipherment

Everything documented at
<http://stp.lingfil.uu.se/~bea/copiale>
Google: "copiale"

First lawbook
of the ☉ e ☉

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 Λ, 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 Λ, who asks him:

First, if he desires to become ☉.

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

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

The candidate answers yes.

lit:mz||bl
vχz|l̥s̥k̥p̥k̥/wn
πo|p̥h̥Δ̥ēz̥c̥ā=ʒ̥l̥ūb̥☉̥u̥r̥☉̥z̥

δ̥h̥z̥i+δ̥i̥n̥p̥l̥h̥i̥n̥c̥f̥.

c̥ū|f̥ēz̥t̥p̥t̥g̥h̥l̥:k̥

δ̥x̥ū̥h̥ē̥η̥+χ̥r̥k̥m̥l̥i̥z̥:ȳ̥ʒ̥l̥δ̥o̥i̥q̥z̥i̥l̥x̥ā|ē̥c̥:u̥δ̥.

f̥ū̥r̥f̥o̥i̥k̥l̥l̥i̥=̥c̥z̥.

m̥ʒ̥h̥r̥ḁ̄+Δ̥g̥ȳ̥x̥z̥δ̥m̥n̥k̥f̥h̥h̥t̥+̥i̥f̥.

k̥m̥ū̥r̥:ʒ̥z̥i̥δ̥f̥|ȳ̥ʒ̥h̥ē̥j̥h̥z̥l̥n̥π̥ḁ̄z̥b̥Δ̥g̥z̥=ʒ̥l̥z̥u̥p̥q̥c̥l̥ḁ̄r̥g̥k̥l̥h̥

h̥r̥h̥h̥l̥z̥i̥n̥p̥l̥|Δ̥r̥δ̥h̥l̥ḁ=g̥z̥w̥π̥ȳ̥ē̥c̥Δ̥r̥δ̥Δ̥+b̥z̥h̥r̥i̥x̥ȳ̥i̥n̥z̥u̥f̥l̥z̥

π̥k̥i̥ʒ̥δ̥r̥=̥f̥|ū̥l̥s̥k̥m̥δ̥m̥|z̥h̥g̥ḁ̄|k̥h̥=̥l̥h̥|l̥x̥δ̥w̥f̥:π̥l̥b̥i̥f̥u̥m̥ȳ̥i̥z̥v̥

z̥ḁ̄i̥x̥ʒ̥r̥m̥π̥z̥h̥l̥c̥δ̥o̥g̥g̥z̥ū̥+ʒ̥k̥h̥n̥x̥ē̥z̥g̥h̥l̥h̥h̥i̥h̥h̥l̥i̥z̥t̥n̥k̥i̥r̥δ̥ȳ̥

m̥ḁ̄+h̥h̥r̥z̥δ̥z̥n̥b̥s̥h̥t̥:χ̥r̥k̥r̥p̥δ̥h̥Δ̥c̥ū̥l̥g̥=̥n̥z̥k̥p̥x̥o̥on̥z̥h̥ʒ̥f̥h̥n̥π̥

z̥ē̥ȳ̥n̥g̥=̥r̥π̥g̥δ̥Δ̥z̥n̥z̥h̥|k̥π̥h̥i̥x̥ȳ̥r̥i̥g̥π̥u̥Λ̥b̥g̥l̥i̥f̥b̥δ̥ū̥f̥h̥h̥z̥δ̥l̥e̥

z̥h̥δ̥|ū̥r̥c̥f̥z̥q̥δ̥l̥b̥h̥h̥m̥δ̥:

n̥i̥r̥f̥c̥i̥r̥e̥o̥p̥h̥δ̥r̥δ̥p̥z̥δ̥ū̥h̥z̥ḁ̄k̥☉̥g̥s̥=̥δ̥m̥ē̥i̥z̥u̥l̥i̥

g̥s̥m̥ū̥o̥o̥l̥δ̥h̥|o̥π̥ē̥g̥u̥δ̥δ̥z̥r̥Δ̥i̥z̥g̥k̥r̥x̥u̥g̥π̥ū̥z̥k̥☉̥h̥i̥h̥ʒ̥b̥=̥n̥

l̥ē̥r̥m̥δ̥z̥f̥:u̥ḁ=̥r̥z̥l̥Δ̥h̥r̥h̥δ̥m̥h̥ḁ̄z̥δ̥i̥|δ̥i̥r̥f̥i̥δ̥δ̥ȳ̥l̥o̥z̥i̥ē̥g̥c̥ū̥h̥

r̥s̥ē̥h̥l̥

c̥r̥=̥f̥|f̥ḁ̄h̥z̥b̥m̥Δ̥c̥:z̥δ̥.

h̥z̥i̥h̥l̥:δ̥g̥l̥e̥z̥h̥ḁ̄z̥Δ̥n̥π̥ḁ̄z̥☉̥p̥s̥k̥r̥δ̥i̥z̥t̥m̥ḁ̄ȳ̥δ̥u̥l̥i̥=̥n̥π̥

z̥r̥s̥=̥n̥h̥k̥f̥r̥z̥r̥i̥n̥δ̥z̥i̥p̥g̥π̥c̥ȳ̥ʒ̥h̥f̥δ̥b̥i̥l̥ʒ̥k̥n̥g̥h̥z̥ȳ̥t̥i̥δ̥c̥s̥=̥b̥+̥n̥ʒ̥u̥δ̥

x̥δ̥i̥r̥:ū̥l̥v̥i̥o̥on̥.

l̥π̥δ̥z̥f̥ʒ̥h̥g̥z̥ȳ̥π̥r̥l̥n̥m̥l̥m̥Δ̥r̥l̥h̥l̥o̥e̥q̥n̥r̥.

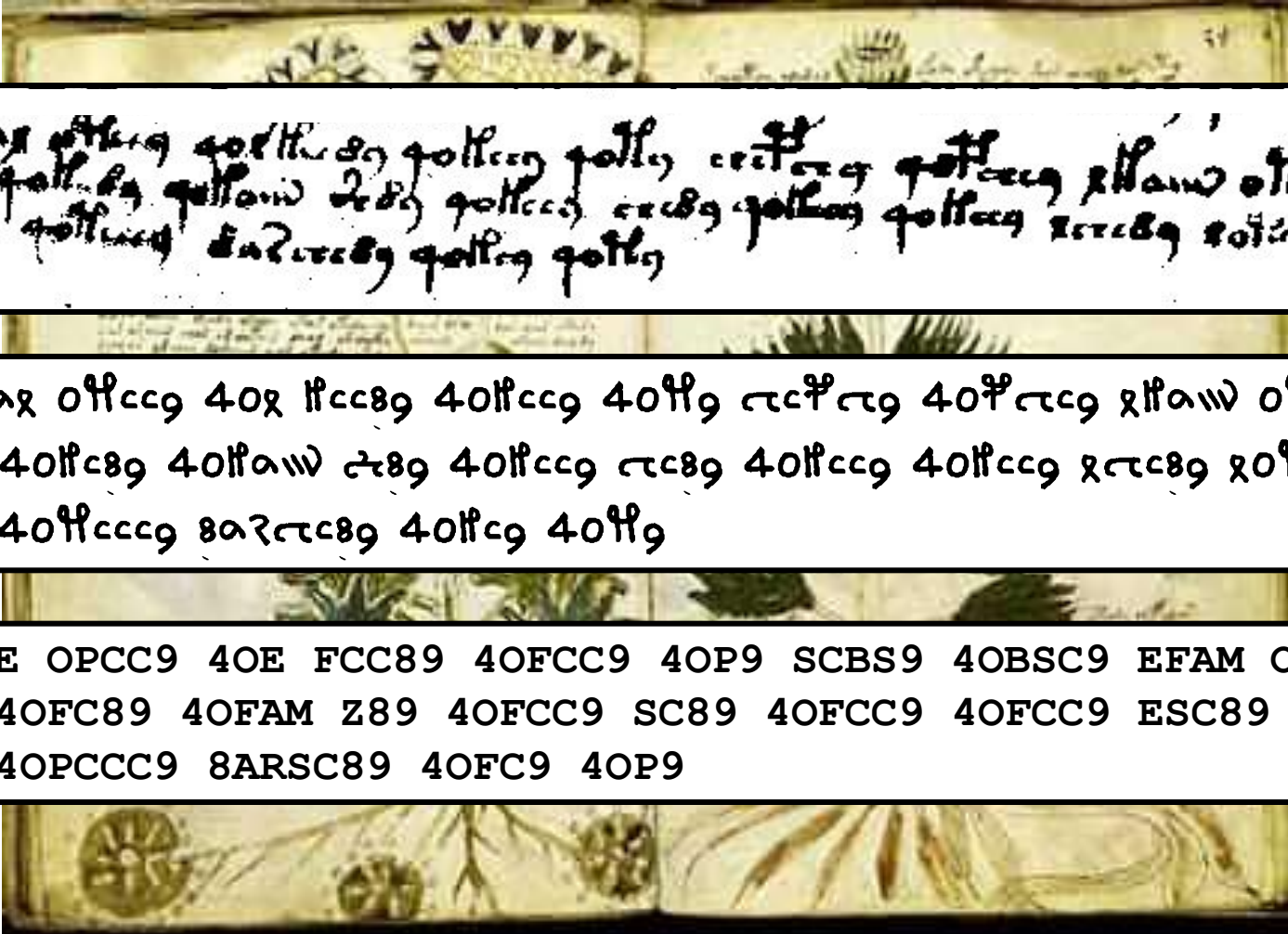
Voynich Manuscript



- Medieval illustrated manuscript
- 235 pages, 6 sections, 38k word tokens, 35 letter types
- Undeciphered



Voynich Manuscript



Handwritten text in Voynich script, likely from a page in the Voynich Manuscript. The text is written in a cursive style and appears to be a list or a series of entries, possibly related to the botanical illustrations on the adjacent page.

Handwritten text in Voynich script, likely from a page in the Voynich Manuscript. The text is written in a cursive style and appears to be a list or a series of entries, possibly related to the botanical illustrations on the adjacent page.

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

Voynich Letter Substitution

Latin letter trigram model

quo_vade_bre...

a → {all Voynich letters}
b → {all Voynich letters}
c → {all Voynich letters}
...
z → {all Voynich letters}
_ → _

V A S 9 2 _ 9 F A E _ A R _ A P A M _ ...

| Input | Decipherment |
|---------------|---------------|
| VAS92 9FAE AR | quiss squm is |
| APAM ZOE ZOR9 | onum pom |
| QOR92 9 FOR | quiss hates s |
| ZOE89 ... | qum hatis ... |



Letter Clustering

Trigram model over {a, b, _}

a a _ b a b _ a b a a _ ...



i n _ t h e _ t o w n _ w h e r e _ i _ w a s ...

Sample tagging with learned model:

a b _ b b a _ b a b b _
i n _ t h e _ t o w n _

b b a b a _ a _ ...
w h e r e _ i _ ...

Letter Clustering

Trigram model over {a, b, _}

a a _ b a b _ a b a a _ ...

a → {all Voynich letters}

b → {all Voynich letters}

_ → _

V A S 9 2 _ 9 F A E _ A R _ A P A M _ ...

Sample tagging with learned model:

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|-----|
| ? | ? | ? | ? | ? | _ | ? | ? | ? | ? | _ | ? | ? | _ | |
| V | A | S | 9 | 2 | _ | 9 | F | A | E | _ | A | R | _ | |
| ? | ? | ? | ? | _ | ? | ? | ? | _ | ? | ? | ? | ? | _ | ... |
| A | P | A | M | _ | Z | O | E | _ | Z | O | R | 9 | _ | ... |

Letter Clustering

Trigram model over {a, b, _}

a a _ b a b _ a b a a _ ...



V A S 9 2 _ 9 F A E _ A R _ A P A M _ ...

Sample tagging with learned model:

b b b b a _ a b b a _ b a _
V A S 9 2 _ 9 F A E _ A R _

b b b a _ b b a _ b b b a _ ...
A P A M _ Z O E _ Z O R 9 _ ...

Letter Clustering

$P(\text{letter} \mid \text{tag})$

English

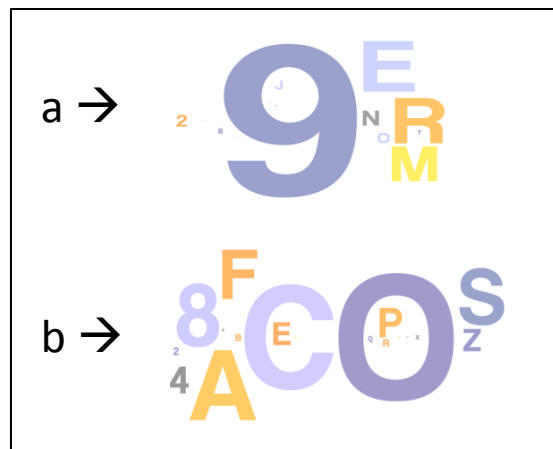


$P(\text{tag} \mid \text{letter})$

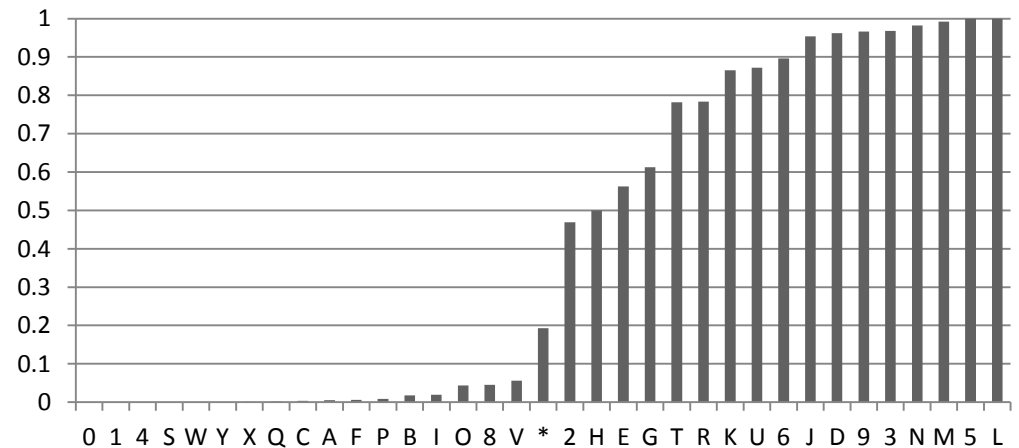
$P(a)$



Voynich



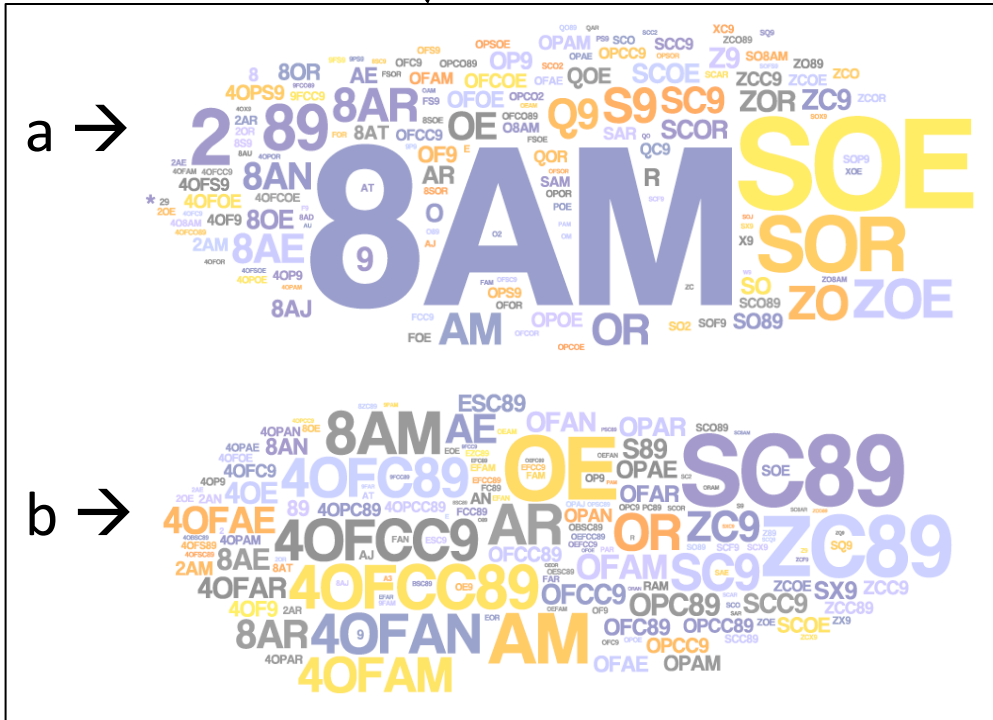
$P(a)$



Word Clustering

Bigram model over $\{a, b\}$

a a b a b a b a a ...



VAS92 9FAE AR APAM ZOE ZOR9 QRC2 9 ...

Sample tagging with learned model:

a a a a a

VAS92 9FAE AR APAM ZOE

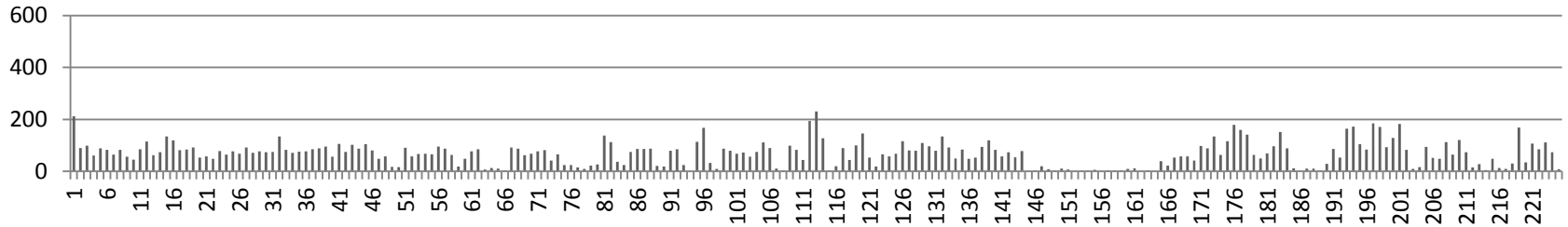
| | | | | | |
|------|------|---|-----|-------|-----|
| a | a | a | a | a | ... |
| ZOR9 | QRC2 | 9 | FOR | ZOE89 | ... |



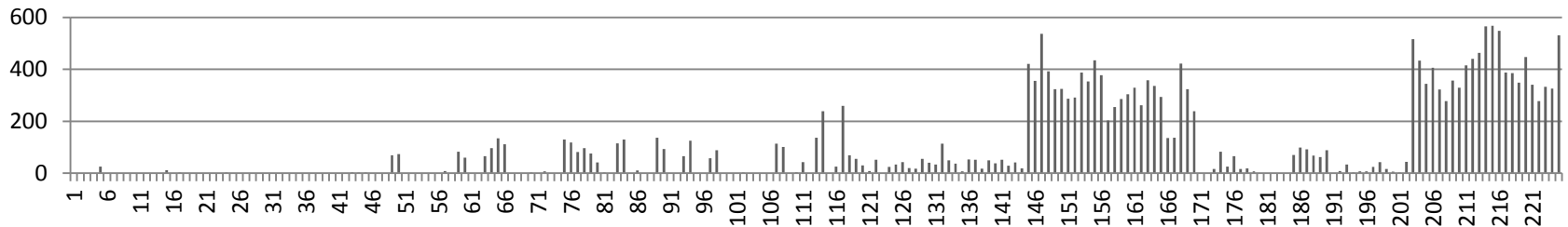
Word Clustering

Voynich words tagged as “a”

← pages →



Voynich words tagged as “b”

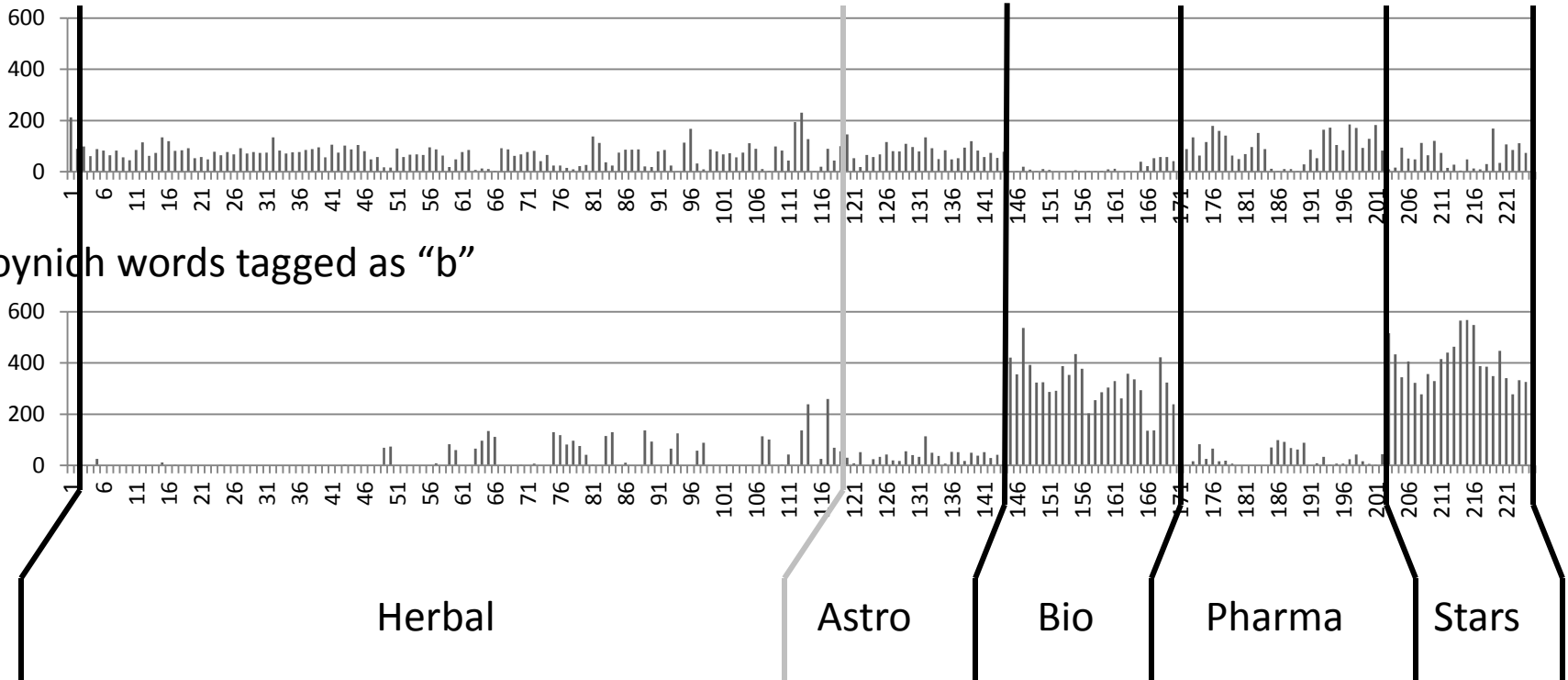


Word Clustering

Voynich words tagged as “a”

← pages →

Voynich words tagged as “b”



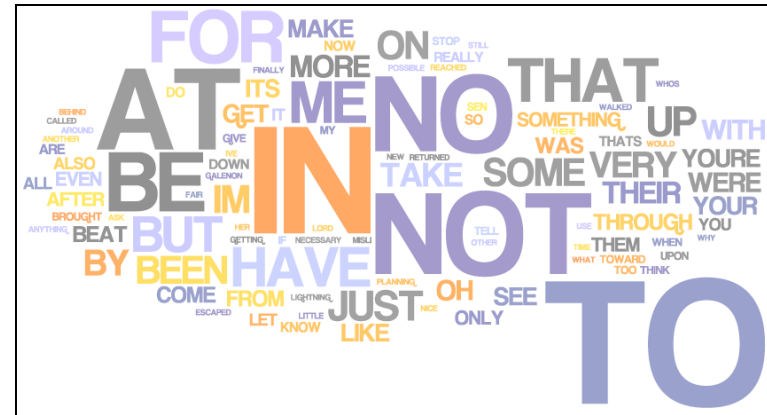
Voynich sections, per drawings observed.
Captain Currier's "two languages" (1976).

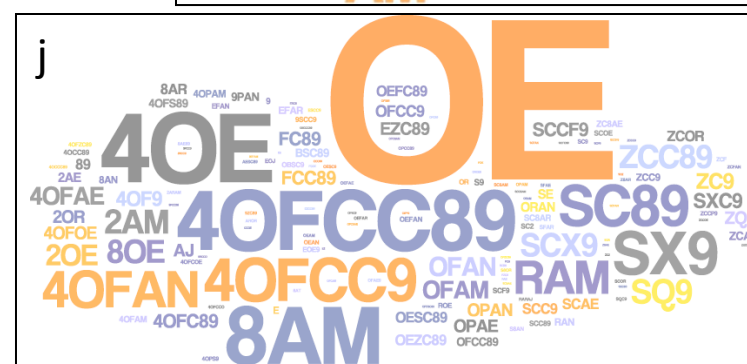
A word cloud visualization of the sentence "My attitude towards her and his". The words are arranged in a grid-like pattern, with each word occupying a space roughly proportional to its frequency in the sentence. The words are: "MY", "ATTITUDE", "TOWARDS", "HER", "AND", "HIS". The words "MY", "ATTITUDE", "TOWARDS", "HER", "AND", "HIS" are in a larger font size, while "TOWARDS" is in a smaller font size. The words are colored in various shades of blue, green, and yellow.

etc



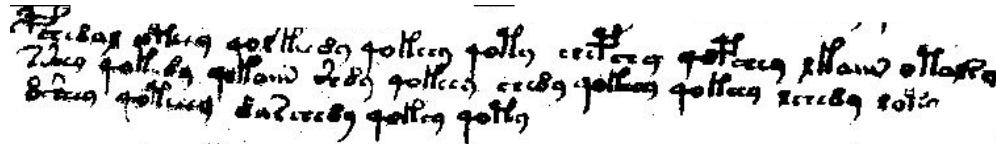
etc





Other Unsolved Ciphers

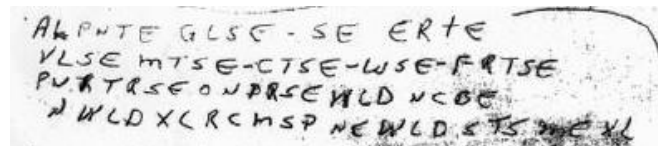
Voynich Manuscript (1400) Reddy & Knight 11



Zodiac 408 Serial Killer (1967)



FBI cipher (1999)



Kryptos (1990)

OBKR
UOXOGHULBSOLIFBBWFLRVQQPRNGKSSO
TWTQSJQSSEKZZWATJKLUDIAWINFBNYP
VTTMZFPKWGDKZXTJCDIGKUHUAUEKCAR

Lost
Language
Decipherment



Snyder, Barzilay, Knight 10

Unsupervised
Training for NLP



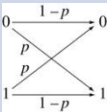

Machine
Translation ...

Plan for This Talk



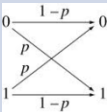

- Break a series of codes
 - Simple letter substitution
 - Phonetic substitution
 - archaeology
 - transliteration
 - Word substitution
 - Foreign language as cipher
- Bonus
 - Two historical ciphers
 - Final thought on translation and cryptography



Future Prospects for Translation

| | | Cryptography | Translation |
|-----------------------------------------|------------------------------------------------------------------------------------|-----------------------------------------------------------|-------------------|
| Manual |  | Manual encoding | Human translation |
| Mechanical |  | 1920s Mechanical encoding; intuition-based decryption | |
| Mathematical |  | 1950s Computer decryption, based on information theory | |
| Higher math, deeper understanding |  | 1980s Public-key systems, based on number theory | |

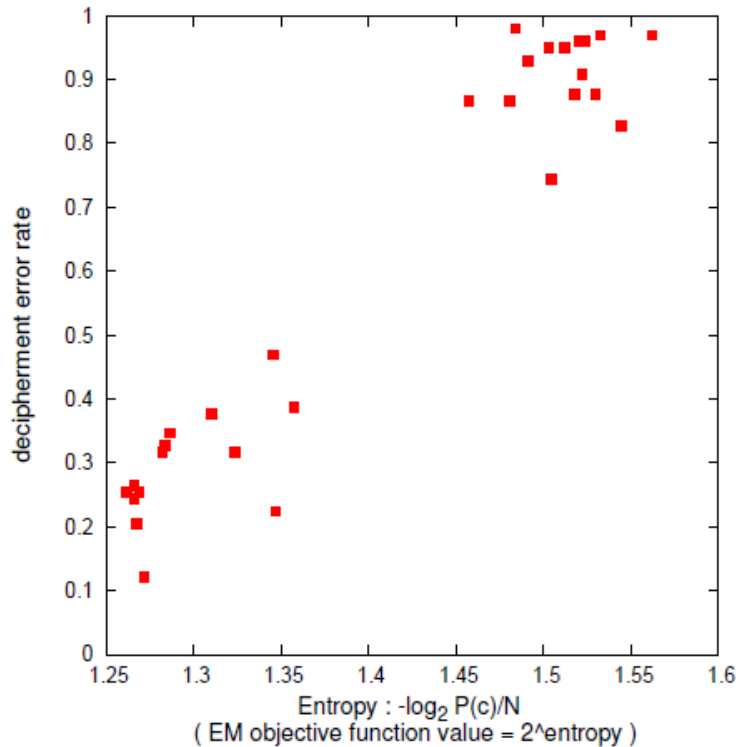
Future Prospects for Translation

| | | Cryptography | Translation |
|-----------------------------------|------------------------------------------------------------------------------------|--------------------------------------------------------|----------------------|
| Manual |  | Manual encoding | Human translation |
| Mechanical |  | 1920s Mechanical encoding; intuition-based decryption | 1960s Rule-based MT |
| Mathematical |  | 1950s Computer decryption, based on information theory | 1990s Statistical MT |
| Higher math, deeper understanding |  | 1980s Public-key systems, based on number theory | 2020s ??? ??? ??? |

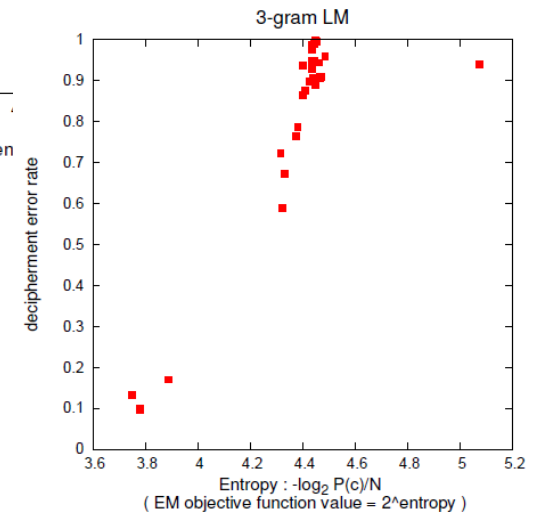
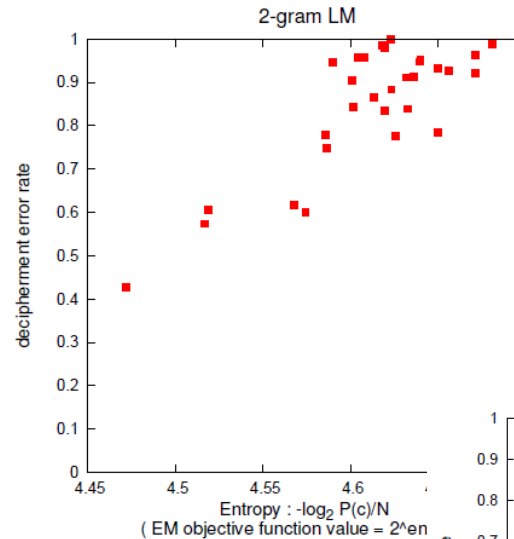
thanks

Random Restarts are Critical

English 98-letter cipher, 3-gram LM

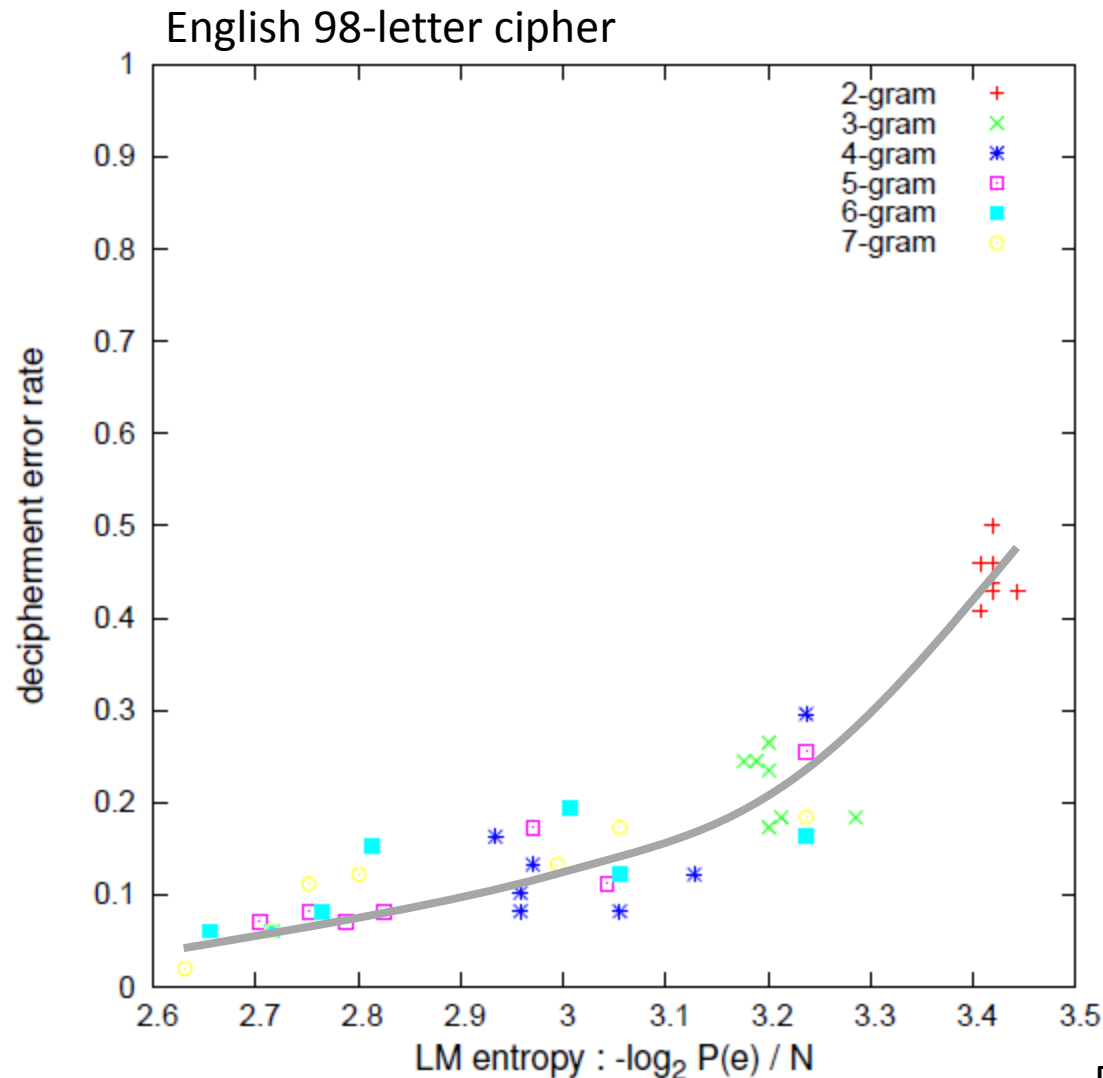


Japanese
syllable
cipher



even people do restarts!

Good Language Models are Critical



[Ravi & Knight 09b]

Deterministic Substitution Constraint

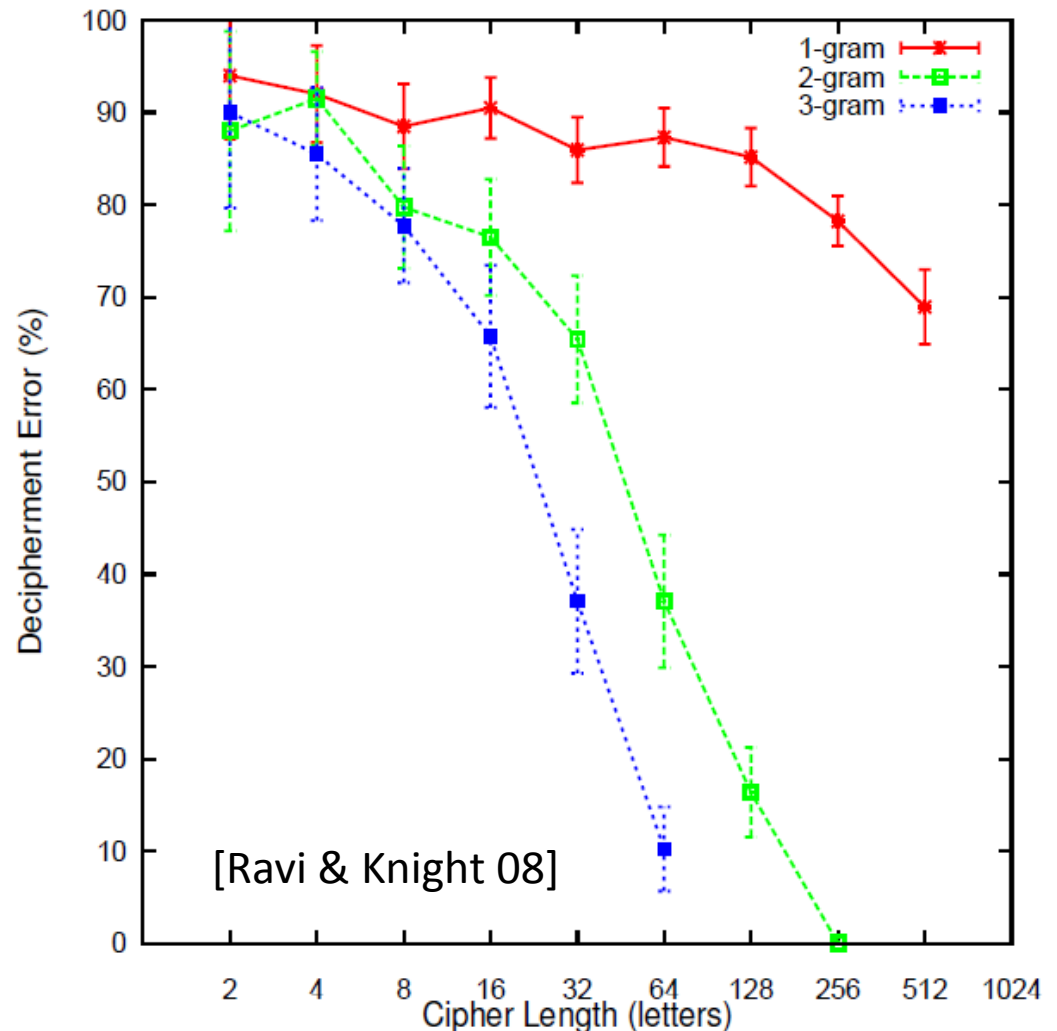
Using ILP instead of EM

- * Search only over deterministic keys.
- * Exact, no restarts.



| Cipher Length | EM error | ILP error |
|---------------|----------|--------------|
| 52 | 85 % | 21 % |
| 98 | 45 % | 12 % |
| 414 | 10 % | 0.5 % |

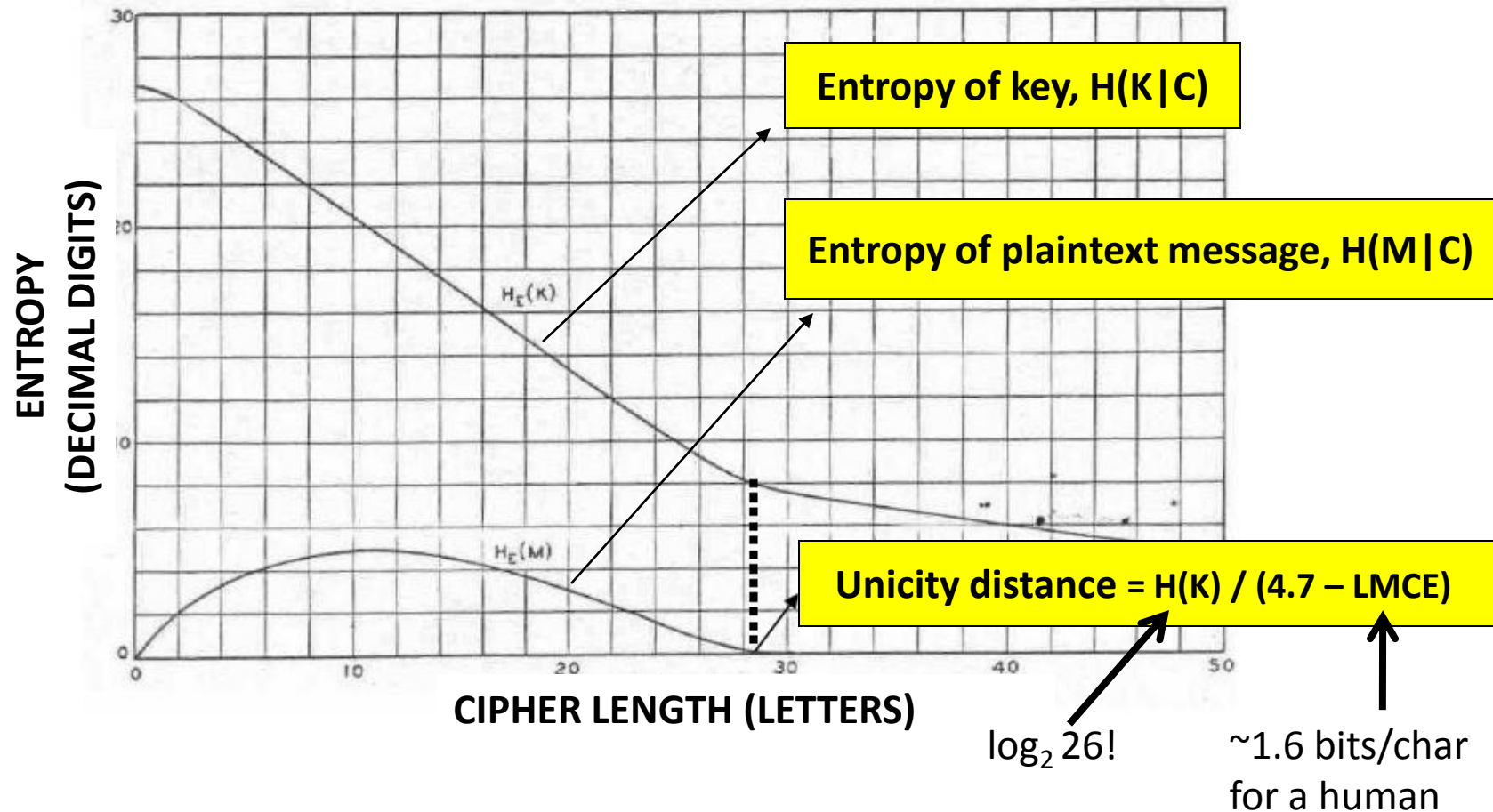
Using 2-gram letter-based LM



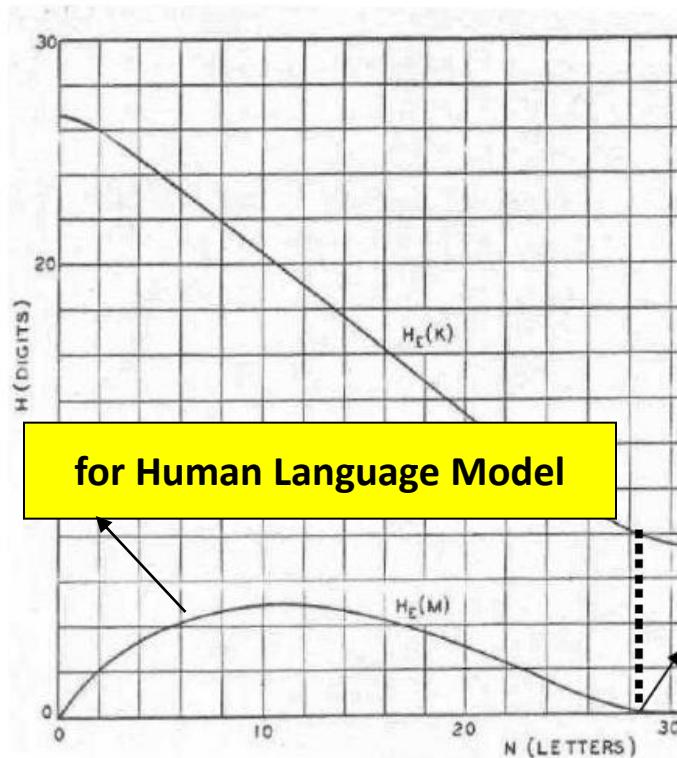
[Shannon 46, 49]

“Communication Theory of Secrecy Systems”

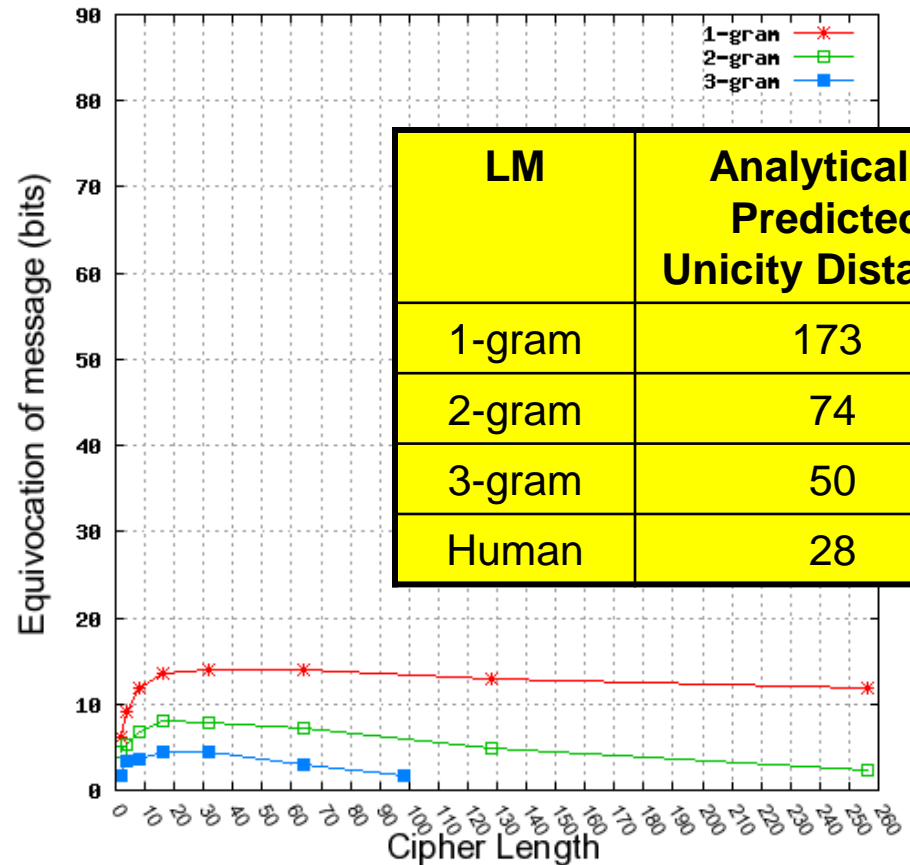
- Shannon analytically predicted uncertainty about key and message
- Graphed it for a human-level language model



Verifying Shannon's Prediction of Plaintext Message Uncertainty



ANALYTIC CURVES
(Shannon's)

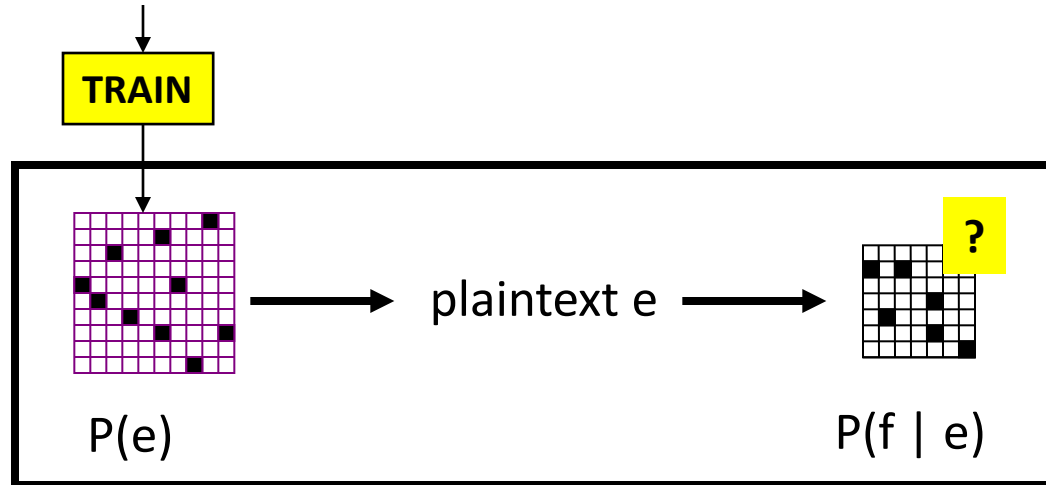


ACTUAL CURVES
(ours)

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 يتعين 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 يتعين 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".

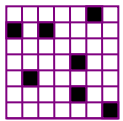
الخيار بكل ترحاب من جانبه شجب رئيس الحكومة الفلسطينية أحمد قريع الطابع الأحادي الجانب للانسحاب الإسرائيلي من غزة، وأكد أن إسرائيل تريد مغادرة هذه الأراضي لتعزيز سيطرتها على الضفة الغربية وقال قريع في كلمة له خلال مؤتمر نظمته وزارة الأوقاف في رام الله "سينسحبون من غزة ولكننا لا نعرف ما هو شكل هذا الانسحاب وماذا سيتركون، وما هو مصير المعابر والحدود، وكل ذلك غامض لأنه قرار أحادي الجانب

Word Substitution Cipher

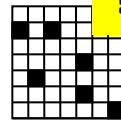
.....France.....Britain.....Canada...
Mexico.....Indonesia.....Malaysia...
Britain.....Canada.....Australia...
Britain.....France.....Indonesia.....
Mexico.....Australia.....France...
 ...Britain.....

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

TRAIN



plaintext e



$P(\text{sentence has } w1 \mid \text{sentence has } w2)$

$P(f \mid e) =$
7 x 7 subst table

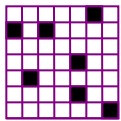
.....knd!.....bryT!ny!
knd!.....
 !lmksyk.....
 ...!ndwnysy!.....!lmksyk.....
bryT!ny!.....!m!lyzy!...
bryT!ny!.....frns!.....
!str!ly!.....!ndwnysy!...
frns!.....frns!
frns!.....bryT!ny!.....
!str!ly!.....

Word Substitution Cipher

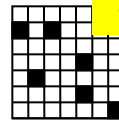
.....France.....Britain.....Canada...
Mexico.....Indonesia.....Malaysia...
Britain.....Canada.....Australia...
Britain.....France.....Indonesia.....
Mexico.....Australia.....France...
 ...Britain.....

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

TRAIN



plaintext e



$P(\text{sentence has } w1 \mid \text{sentence has } w2)$

$P(f \mid e) =$
7 x 7 subst table

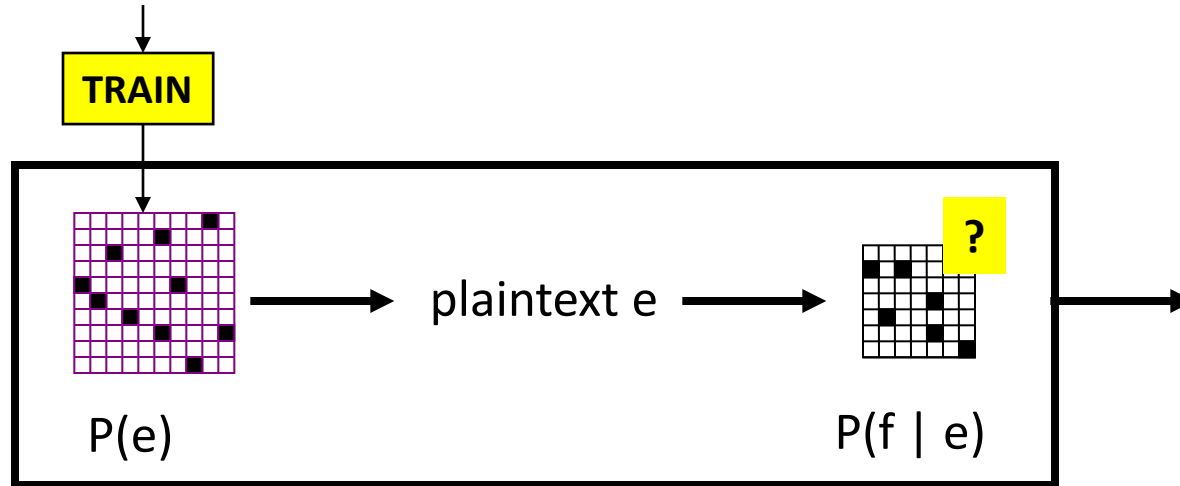
.....knd!.....bryT!ny!
knd!.....
 !lmksyk.....
 ...!ndwnysy!.....!lmksyk.....
bryT!ny!.....m!lyzy!...
bryT!ny!.....frns!.....
!str!ly!.....!ndwnysy!...
frns!.....frns!
frns!.....bryT!ny!.....
!str!ly!.....

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

[Knight et al 06]

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 يتعين 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 يتعين 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 جانيه شجب رئيس الحكومة الفلسطينية أحمد قريع الطابع الأحادي الجانب للانسحاب الإسرائيلي من غزة، وأكد أن إسرائيل تريد مغادرة هذه الأراضي 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 الجانب

Zodiac Killer Ciphers

Zodiac 408 (solved, 1969)

Δ ▣ P / Z / U B ▣ X O R π 9 X π B
W V + ε G Y F O Δ H P ▣ K π ρ Y ε
M J Y Λ U I X Δ ρ T ⊥ N Q Y D ● ρ
S ϕ / Δ ▣ B P O R A U ▣ ε R J ρ E
X Λ L M Z J O R \ 9 F H V W ε Δ Y
▣ + ρ G D Δ K I ● ρ X Δ ● ϕ S ϕ
R N ⊥ I Y E J O Δ ρ G B T Q S ▣ B
L O / P ▣ B ▣ X ρ E H M U Λ R R X

Ϸ Z K ρ 9 I ● W ρ I Δ ● L M ρ Δ ▣
B P D R + τ π ρ \ N ϕ ε E U H K F
Z Ϸ 9 O V W I ● + ⊥ L ● J Λ R O H
I Δ D R ▣ T Y ρ \ ρ ε / ▣ X J Q A
P ● M Δ R U ⊥ ▣ L ● N V E K H π G
ρ I I J X ● Δ Δ L M J N A ● Z ϕ P
ϕ U 9 X A Δ ▣ B V W \ + V T ⊥ O P
Λ π S ρ J ϕ U ε O Δ ρ ϕ G ▣ ▣ I M
N K ● S Ϸ E / Δ ▣ ▣ Z Ϸ A P ▣ B V
9 ε X ρ W ρ ▣ F ▣ Δ Ϸ + ▣ Δ A Δ B
▣ O T ● R U Ϸ + ▣ ρ Y ρ ▣ λ S ρ W
V Z ε G Y K E ρ T Y A Δ ▣ ▣ L ⊥ ▣
H I F B X Δ ϕ X A D ρ \ Δ L I π ρ
▣ ε D ▣ ▣ ρ ε ρ P O R X Q F ▣ G Ϸ
Z ▣ J T ⊥ ρ ▣ Δ J I + ρ B P Q W O
V E X ρ Δ W I ρ ρ E H M ρ π U I K



Zodiac 340 (still unsolved)

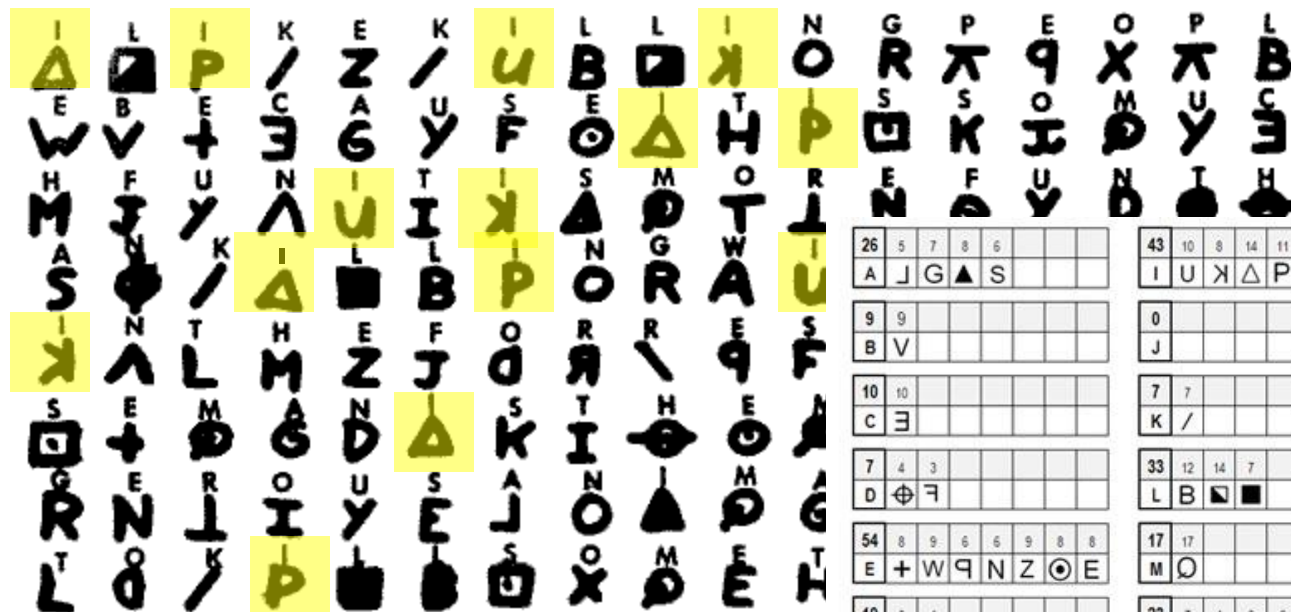
HER > 9 J Λ V P X I ● L T G ● Q
N 9 + B ϕ ▣ O ▣ D W Y · < ▣ K Ϸ ρ
B X π Ϸ M + u z G W ϕ ϕ L ▣ ϕ H J
S 9 9 Δ Λ J Δ ▣ V O 9 O + + R K ●
▣ Δ M + ϕ ⊥ T Q I ● F P + ρ X /
9 Δ R Λ F J O - ▣ ρ C Ϸ F > ● ρ ϕ
▣ ● + K ρ ▣ π ● u Ϸ X G V · ϕ L I
ϕ G ● J Ϸ T ▣ O + ▣ N Y ϕ + ▣ L Δ
O < M + 8 + Z R ● F B Ϸ Y A O ● K
- ϕ J u v + Λ J + O 9 Δ < F B X -
U + R / ● ⊥ E I D Y B 9 8 T M K O
● < Ϸ J R J I ▣ ● T ● M · + P B F
ϕ ϕ Δ S Y ▣ + N I ● F B Ϸ ϕ π Δ R
J G F N Λ Ϸ ● ● ● 8 · Ϸ V ● ⊥ + +
Y B X ● ▣ π ● Δ C E > V U Z ● - +
I Ϸ · O ϕ B K ϕ O 9 Λ · Ϸ M ρ G ●
R Ϸ T + L ● ρ C < + F J W B I ● L
+ + ϕ W C ϕ W Ϸ P O S H T / ϕ ϕ 9
I F X ρ W < Δ ⊥ B ρ Y O B ▣ - Ϸ Ϸ
> M D H N 9 X S ϕ Z O Δ A I K π +

COOP-SFPO
1596-78
7-14-78 GVL
7-22-78

#2 H-9-67

Zodiac Serial Killer

408-letter cipher (solved):



(plus two more sections)

| | | | | | |
|----|---|---|---|---|--|
| 26 | 5 | 7 | 8 | 6 | |
| A | J | G | ▲ | S | |

| | | | | | |
|---|---|--|--|--|--|
| 9 | 9 | | | | |
| B | V | | | | |

| | | | | | |
|----|----|--|--|--|--|
| 10 | 10 | | | | |
| C | E | | | | |

| | | | | | |
|---|---|---|--|--|--|
| 7 | 4 | 3 | | | |
| D | ⊕ | F | | | |

| | | | | | | | |
|----|---|---|---|---|---|---|---|
| 54 | 8 | 9 | 6 | 6 | 9 | 8 | 8 |
| E | + | W | 9 | N | Z | ⊙ | E |

| | | | | | | | |
|----|---|---|--|--|--|--|--|
| 10 | 6 | 4 | | | | | |
| F | J | Q | | | | | |

| | | | | | | | |
|----|----|--|--|--|--|--|--|
| 11 | 11 | | | | | | |
| G | R | | | | | | |

| | | | | | | | |
|----|---|---|--|--|--|--|--|
| 16 | 8 | 8 | | | | | |
| H | M | ⊖ | | | | | |

| | | | | |
|----|----|---|----|----|
| 43 | 10 | 8 | 14 | 11 |
| I | U | X | △ | P |

| | | | | |
|---|--|--|--|--|
| 0 | | | | |
| J | | | | |

| | | | | | |
|---|---|--|--|--|--|
| 7 | 7 | | | | |
| K | / | | | | |

| | | | | |
|----|----|----|---|--|
| 33 | 12 | 14 | 7 | |
| L | B | ■ | ■ | |

| | | | | | |
|----|----|--|--|--|--|
| 17 | 17 | | | | |
| M | ○ | | | | |

| | | | | |
|----|---|---|---|---|
| 23 | 7 | 4 | 6 | 6 |
| N | ○ | Φ | Λ | D |

| | | | | |
|----|---|---|---|---|
| 28 | 6 | 7 | 9 | 6 |
| o | □ | T | X | J |

| | | | | | |
|---|---|--|--|--|--|
| 7 | 7 | | | | |
| P | ∟ | | | | |

| | | | | | |
|---|--|--|--|--|--|
| 0 | | | | | |
| Q | | | | | |

| | | | | | |
|----|---|---|---|--|--|
| 19 | 7 | 7 | 5 | | |
| R | ⊥ | Я | \ | | |

| | | | | |
|----|---|---|---|---|
| 22 | 7 | 6 | 3 | 6 |
| S | F | K | △ | □ |

| | | | | |
|----|---|---|---|----|
| 33 | 7 | 8 | 8 | 10 |
| T | ● | L | H | I |

| | | | | | |
|----|----|--|--|--|--|
| 10 | 10 | | | | |
| U | Y | | | | |

| | | | | | |
|---|---|--|--|--|--|
| 6 | 6 | | | | |
| V | ∪ | | | | |

| | | | | | |
|---|---|--|--|--|--|
| 8 | 8 | | | | |
| W | A | | | | |

| | | | | | |
|---|---|--|--|--|--|
| 1 | 1 | | | | |
| X | ∟ | | | | |

| | | | | | |
|---|---|--|--|--|--|
| 8 | 8 | | | | |
| Y | □ | | | | |

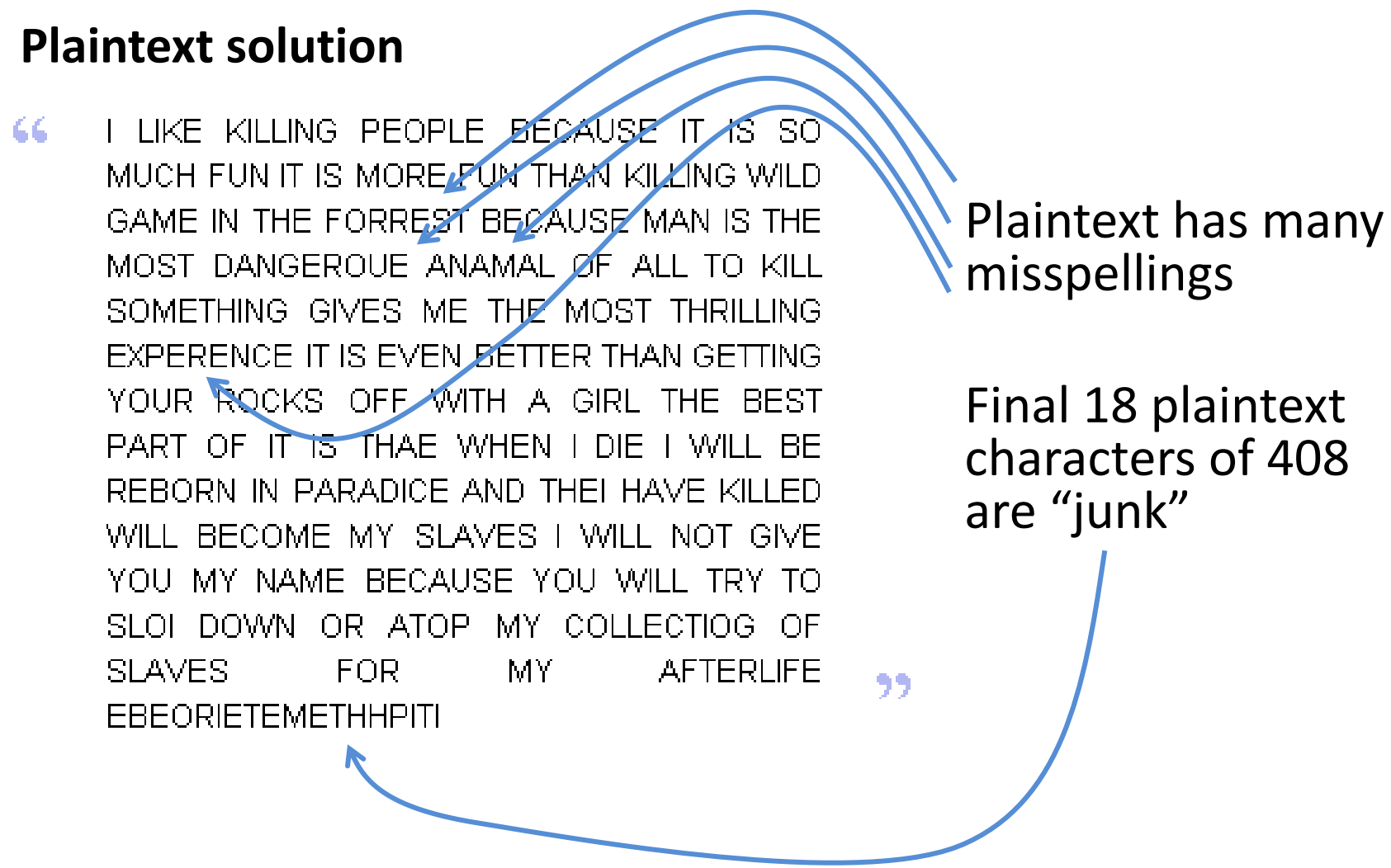
| | | | | | |
|---|--|--|--|--|--|
| 0 | | | | | |
| Z | | | | | |

| | | | | | |
|-----|--|--|--|--|--|
| 408 | | | | | |
|-----|--|--|--|--|--|

Zodiac Serial Killer

Plaintext solution

“ I LIKE KILLING PEOPLE BECAUSE IT IS SO
MUCH FUN IT IS MORE FUN THAN KILLING WILD
GAME IN THE FORREST BECAUSE MAN IS THE
MOST DANGEROUE ANAMAL OF ALL TO KILL
SOMETHING GIVES ME THE MOST THRILLING
EXPERENCE IT IS EVEN BETTER THAN GETTING
YOUR ROCKS OFF WITH A GIRL THE BEST
PART OF IT IS THAE WHEN I DIE I WILL BE
REBORN IN PARADICE AND THEI HAVE KILLED
WILL BECOME MY SLAVES I WILL NOT GIVE
YOU MY NAME BECAUSE YOU WILL TRY TO
SLOI DOWN OR ATOP MY COLLECTIOG OF
SLAVES FOR MY AFTERLIFE ”
EBEOR IETEMETHHPITI



Plaintext has many misspellings

Final 18 plaintext characters of 408 are “junk”

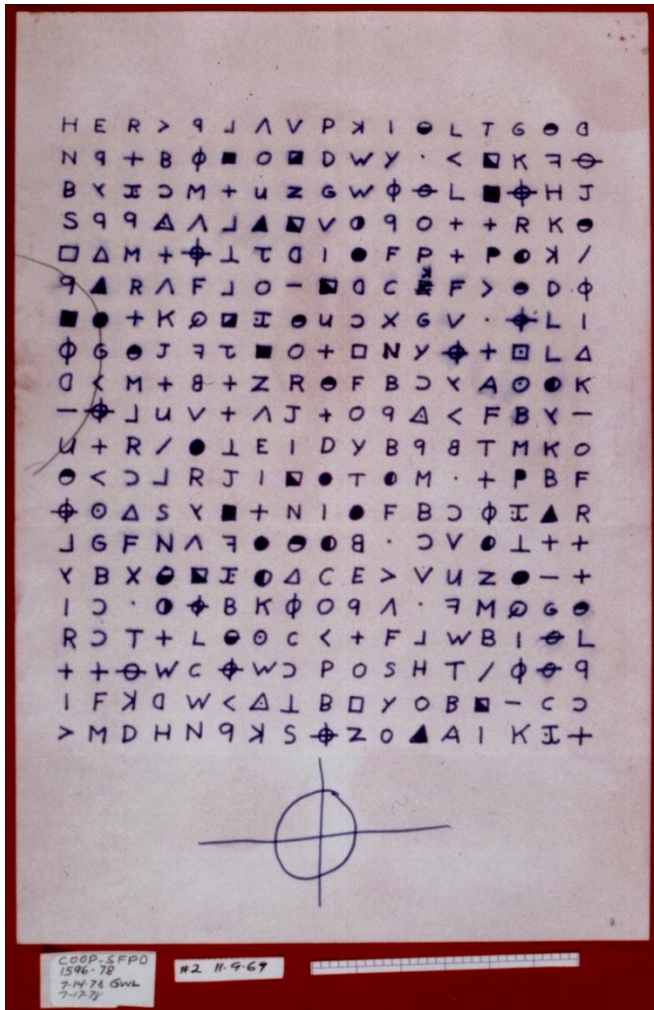
Deciphering Zodiac 408

Bayesian models

Extended Carmel finite-state toolkit to do Bayesian inference.
[Chiang et al 10]

| Language Model | Initial Sample | Decipherment Error |
|--------------------------------------------|-----------------|--------------------|
| 3-gram | Random | 62.3 / 48.5 / 47.4 |
| 5-gram | Random | all wrong! |
| “ | 3-gram solution | 42.6 |
| Word 1-gram | Random | all wrong! |
| <i>Interpolated</i> 5-gram and word 1-gram | Random | 79.2 |
| “ | 5-gram solution | 3.3 / 2.6 |

Unsolved Zodiac 340



Has no obvious reading order bias:

| % cipher bigram types that repeat (freq > 1) | Left/Right order | Up/Down order | Diag. North-East | Diag. South-East |
|----------------------------------------------|------------------|---------------|------------------|------------------|
| Zodiac 408 (solved) | 13 % | 5 | 7 | 5 |
| Zodiac 340 (unsolved) | 7 | 6 | 8 | 5 |

Could be nonsense ... or maybe bigrams are smoothed out via more careful substitutions.