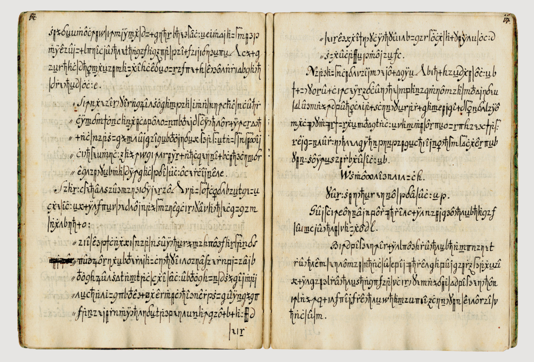
Codebreaking for Traditional Cipher Systems

Davide Dispenza



Overview

Introduction

Substitution Ciphers

Permutation Ciphers

The Copiale Cipher

Introduction

- **Cryptography**

  ALICE(P,K) $\xrightarrow{C}$ BOB(K) secure
  P: Plaintext
  C: Ciphertext
  K: Key

  Perfectly Secure Encryption $\rightarrow \mathrm{I}(\mathrm{M};\mathrm{C}) = 0$

Introduction

- **Cryptography**

  ALICE(P,K) $\xrightarrow{C}$ BOB(K) secure
  P: Plaintext
  C: Ciphertext
  K: Key

  Perfectly Secure Encryption $\rightarrow \mathrm{I}(\mathrm{M};\mathrm{C}) = 0$

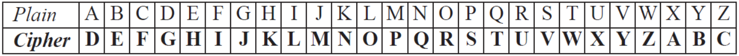
- **Classical Cryptography**
  "Pen and paper" encryption schemes
  Basic elements: Substitution & Permutation

Substitution Ciphers

- Most basic form of encryption
- Every symbol is encoded into another symbol

Caesar Cipher

- Used by Julius Caesar to send military messages
- The alphabet is shifted by some fixed amount
- Example:



Plain	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
Cipher	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**	**N**	**O**	**P**	**Q**	**R**	**S**	**T**	**U**	**V**	**W**	**X**	**Y**	**Z**	**A**	**B**	**C**

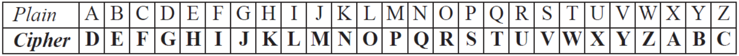
- The message ”HELLO WORLD” becomes ”KHOOR ZRUOG”

Substitution Ciphers

- Most basic form of encryption
- Every symbol is encoded into another symbol

Caesar Cipher

- Used by Julius Caesar to send military messages
- The alphabet is shifted by some fixed amount
- Example:



Plain	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
Cipher	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**	**N**	**O**	**P**	**Q**	**R**	**S**	**T**	**U**	**V**	**W**	**X**	**Y**	**Z**	**A**	**B**	**C**

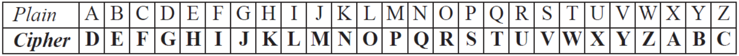
- The message ”HELLO WORLD” becomes ”KHOOR ZRUOG”
- Is it safe?

Substitution Ciphers

- Most basic form of encryption
- Every symbol is encoded into another symbol

Caesar Cipher

- Used by Julius Caesar to send military messages
- The alphabet is shifted by some fixed amount
- Example:



Plain	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
Cipher	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**	**N**	**O**	**P**	**Q**	**R**	**S**	**T**	**U**	**V**	**W**	**X**	**Y**	**Z**	**A**	**B**	**C**

- The message ”HELLO WORLD” becomes ”KHOOR ZRUOG”
- Is it safe?
  Not really, we can easily try all 26 possible combinations

Substitution Ciphers: An Example

- Suppose now we are not just shifting our alphabet
  - We randomly assign a character to another symbol
  - Non-alphabetic characters (including space) can also be used

Substitution Ciphers: An Example

- Suppose now we are not just shifting our alphabet
  - We randomly assign a character to another symbol
  - Non-alphabetic characters (including space) can also be used

Sample of Ciphertext

```
RVRZF19;:-P:8OP-8RHP8:PL1P19RP-LYY8_DP19RZRP;HPLPZL
;YOLFPH_RRWP;:P19RPH1;ZZ;:-P8EP19RPS;:WCP:81PYRHHP
19L:P;:P19RPS8VRSR:1P8EP19RP78W;RHP8EPSR:DP19RPH8N;LY
PL:WP_8Y;1;NLYP_LHH;8:HP9LVRPLNMI;ZRWPHIN9P;:1R:H;
1FCPL:WP7RR:PH8PO;WRYFPW;EEIHRWCP19L1P19R;ZP;:RV;1L7
YRPZRHIY1HPLZRPLYS8H1P;SSRW;L1RYFP_Z8WINRWDP19RP_R
Z;8WP8EPHRRWA1;SRPL:WP9LZVRH1P9LHP7RN8SRPLHPH98Z1P
;:P_8Y;1;NLYPLHP;1P;HP;:PL-Z;NIY1IZLYPYL78IZDPLPH;:-Y
RPFRLZP7Z;:-HP;1HPL__Z8_Z;L1RPEZI;1HP18PSL1IZ;1FP
```

Whole text has 1874 characters

Substitution Ciphers: An Example

Most recurring symbols in text

Symbol	#	%
P	316	16.86%
R	203	10.83%
1	126	6.72%
;	118	6.30%
L	114	6.08%
H	110	5.87%
:	108	5.76%

Most recurring letters in English

Symbol	%
SPACE	19.18%
e	12.70%
t	9.06%
a	8.17%
o	7.51%
i	6.97%
n	6.75%

Substitution Ciphers: An Example

Most recurring symbols in text

Symbol	#	%
P	316	16.86%
R	203	10.83%
1	126	6.72%
;	118	6.30%
L	114	6.08%
H	110	5.87%
:	108	5.76%

Most recurring letters in English

Symbol	%
SPACE	19.18%
e	12.70%
t	9.06%
a	8.17%
o	7.51%
i	6.97%
n	6.75%

Reasonable guess:

P → SPACE
R → e

Substitution Ciphers: An Example

Sample of partly decoded text

eVeZF19;:- :8O -8eH 8: L1 19e -LYY8_D 19eZe ;H L ZL;YOLF
H_eeW ;: 19e H1;ZZ;:- 8E 19e S;:WC :81 YeHH 19L: ;: 19e
S8VeSe:1 8E 19e 78W;eH 8E Se:D 19e H8N;LY L:W _8Y;1;NLY
_LHH;8:H 9LVe LNMI;ZeW HIN9 ;:1e:H;1FC L:W 7ee: H8
O;WeYF W;EEIHeWC 19L1 19e;Z ;:eV;1L7Ye ZeHIY1H LZe
LYS8H1 ;SSeW;L1eYF _Z8WINeWD 19e _eZ;8W 8E HeeWA1;Se
L:W 9LZVeH1 9LH 7eN8Se LH H98Z1 ;: _8Y;1;NLY LH ;1 ;H ;:
L-Z;NIY1IZLY YL78IZD L H;:-Ye FeLZ 7Z;:-H ;1H L__Z8_Z;L1e
EZI;1H 18 SL1IZ;1F ;: 19e S8ZLY LH ;: 19e _9FH;NLY O8ZYWD

Substitution Ciphers: An Example

Sample of partly decoded text

eVeZF19;:- :8O -8eH 8: L1 **19e** -LYY8_D 19eZe ;H L ZL;YOLF H_eeW ;: **19e** H1;ZZ;:- 8E **19e** S;:WC :81 YeHH 19L: ;: **19e** S8VeSe:1 8E **19e** 78W;eH 8E Se:D **19e** H8N;LY L:W _8Y;1;NLY _LHH;8:H 9LVe LNMI;ZeW HIN9 ;:1e:H;1FC L:W 7ee: H8 O;WeYF W;EEIHeWC 19L1 19e;Z ;:eV;1L7Ye ZeHIY1H LZe LYS8H1 ;SSeW;L1eYF _Z8WINeWD **19e** _eZ;8W 8E HeeWA1;Se L:W 9LZVeH1 9LH 7eN8Se LH H98Z1 ;: _8Y;1;NLY LH ;1 ;H ;: L-Z;NIY1IZLY YL78IZD L H;:-Ye FeLZ 7Z;:-H ;1H L__Z8_Z;L1e EZI;1H 18 SL1IZ;1F ;: **19e** S8ZLY LH ;: **19e** _9FH;NLY O8ZYWD

Substitution Ciphers: An Example

Sample of partly decoded text

eVeZF19;:- :8O -8eH 8: L1 **19e** -LYY8_D 19eZe ;H L ZL;YOLF H_eeW ;: **19e** H1;ZZ;:- 8E **19e** S;:WC :81 YeHH 19L: ;: **19e** S8VeSe:1 8E **19e** 78W;eH 8E Se:D **19e** H8N;LY L:W _8Y;1;NLY _LHH;8:H 9LVe LNMI;ZeW HIN9 ;:1e:H;1FC L:W 7ee: H8 O;WeYF W;EEIHeWC 19L1 19e;Z ;:eV;1L7Ye ZeHIY1H LZe LYS8H1 ;SSeW;L1eYF _Z8WINeWD **19e** _eZ;8W 8E HeeWA1;Se L:W 9LZVeH1 9LH 7eN8Se LH H98Z1 ;: _8Y;1;NLY LH ;1 ;H ;: L-Z;NIY1IZLY YL78IZD L H;:-Ye FeLZ 7Z;:-H ;1H L__Z8_Z;L1e EZI;1H 18 SL1IZ;1F ;: **19e** S8ZLY LH ;: **19e** _9FH;NLY O8ZYWD

19e → the

Substitution Ciphers: An Example

- Following the same kind of reasoning, and using more statistical data (words, digrams, trigrams...), we can easily decode the rest of the text

Substitution Ciphers: An Example

- Following the same kind of reasoning, and using more statistical data (words, digrams, trigrams...), we can easily decode the rest of the text

Sample of decoded text

everything now goes on at the gallop. there is a railway speed in the stirring of the mind, not less than in the movement of the bodies of men. the social and political passions have acquired such intensity, and been so widely diffused, that their inevitable results are almost immediately produced. the period of seed-time and harvest has become as short in political as it is in agricultural labour. a single year brings its appropriate fruits to maturity in the moral as in the physical world.

Permutation Ciphers

- More complex than Substitution
  - Given a plaintext with **N** characters
  - Divide it into blocks of length **L**
  - Choose some permutation of **L** elements
  - Apply it to all blocks

Permutation Ciphers

- More complex than Substitution
  - Given a plaintext with **N** characters
  - Divide it into blocks of length **L**
  - Choose some permutation of **L** elements
  - Apply it to all blocks

Example

Plaintext

"IT_IS_RAINING_AND_THE_SKY_IS_GREY" (33 characters)

Divide it into blocks of 3

"IT_ IS_ RAI NIN G_A ND_ THE _SK Y_I S_G REY"

Apply the permutation $1 \rightarrow 3, 3 \rightarrow 1$

"_TI SI_ IAR NIN A_G _DN EHT KS_ I_Y G_S YER"

Ciphertext

"_TISI_IARNINA_G_DNEHTKS_I_YG_SYER"

Permutation Ciphers

- More complex than Substitution
  - Given a plaintext with **N** characters
  - Divide into **m** blocks of length **L**
  - Choose some permutation of **L** elements
  - Apply it to all blocks

Permutation Ciphers

- More complex than Substitution
  - Given a plaintext with **N** characters
  - Divide into **m** blocks of length **L**
  - Choose some permutation of **L** elements
  - Apply it to all blocks

Weaknesses

Permutation Ciphers

- More complex than Substitution
  - Given a plaintext with **N** characters
  - Divide into **m** blocks of length **L**
  - Choose some permutation of **L** elements
  - Apply it to all blocks

Weaknesses

- We know the length of the text → **L** has to be a divisor of **N**
- We can try all the permutations of all block sizes until we start to see real words

Permutation Ciphers

- More complex than Substitution
  - Given a plaintext with **N** characters
  - Divide into **m** blocks of length **L**
  - Choose some permutation of **L** elements
  - Apply it to all blocks

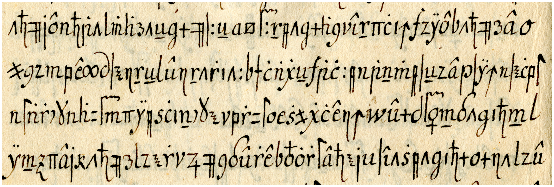
Weaknesses

- We know the length of the text → **L** has to be a divisor of **N**
- We can try all the permutations of all block sizes until we start to see real words
- Can make it harder to break by combining it with substitution

The Copiale Cipher

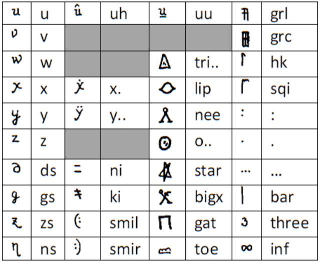
- From 1866
- Discovered in 1970 in an academic archive of East Germany
- Cracked in 2011 by Kevin Knight (USC) and his team
- 75 pages, $\sim 75,000$ characters
- Mix of Roman letters and abstract symbols
- No word Spacing

Sample

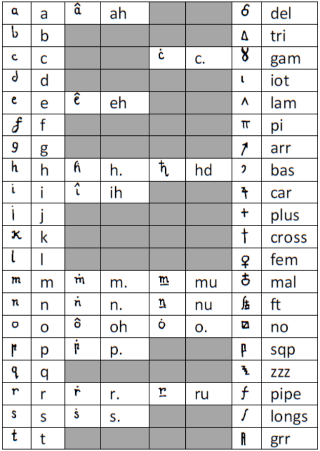


The Copiale Cipher - Transcription

- **First Step**
  All the symbols are transcribed in a machine-readable way



u	u	û	uh	u̲	uu	₶	grl
v	v					▥	grc
w	w			Δ	tri..	ſ	hk
x	x	ẋ	x.	⊖	lip	Γ	sqi
y	y	ÿ	y..	⅄	nee	:	:
z	z			ʘ	o..	·	.
ə	ds	=	ni	☆	star	...	...
g̃	gs	‡	ki	✗	bigx	|	bar
ƺ	zs	(:	smil	Π	gat	3	three
ƞ	ns	:)	smir	⌓	toe	∞	inf

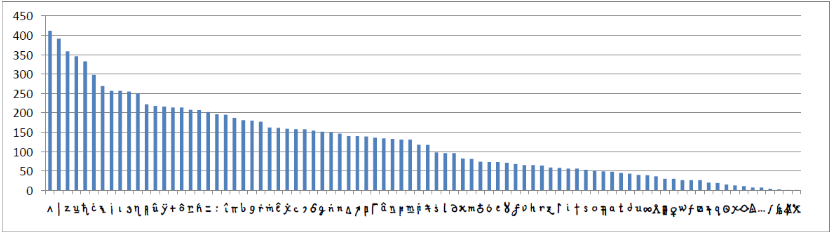


a	a	â	ah			δ	del
b	b					Δ	tri
c	c			ċ	c.	ɣ	gam
d	d					ι	iot
e	e	ê	eh			ʌ	lam
f	f					π	pi
g	g					↗	arr
h	h	ĥ	h.	ɧ	hd	ʔ	bas
i	i	î	ih			‡	car
j	j					+	plus
k	k					†	cross
l	l					♀	fem
m	m	ṁ	m.	m̲	mu	♂	mal
n	n	ṅ	n.	n̲	nu	ʄ	ft
o	o	ô	oh	ȯ	o.	⊠	no
p	p	ṗ	p.			ꝑ	sqp
q	q					ʓ	zzz
r	r	ṙ	r.	r̲	ru	ƒ	pipe
s	s	ṡ	s.			ſ	longs
t	t					₳	grr

The Copiale Cipher - Analysis

- **Second Step**
  Statistical Analysis

Letter Frequencies



Most Common Digrams and Trigrams

ɔ ħ	99	ɔ ħ ∧	47
ċ :	66	ċ : u̲	23
ħ ∧	49	η ɔ ħ	22
: u̲	48	ÿ ɔ ħ	18
z A	44	ɦ ċ |	17

The Copiale Cipher - Clustering

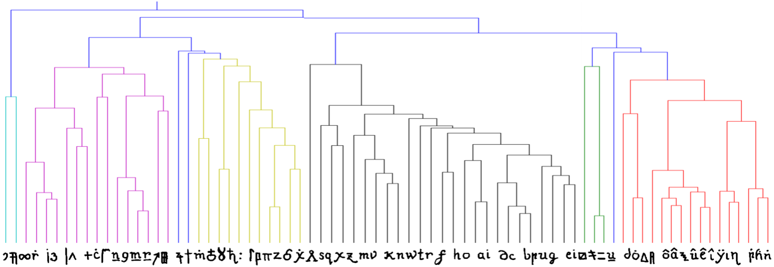
Automatic Clustering of cipher letters

- For each letter **x** a vector is created that captures the distributions of letters that preceed **x**.
- If **x** is preceded 12 times by a, 2 times by b 0 times by c, 3 times by d $\rightarrow [12, 2, 0, 3, ...]$
- Similarly, a vector is created for letters that follow **x**, and the two vectors are concatenated.
- Two letters are similar if the cosine distance $(1 - \cos\theta)$ between the corresponding vectors is small.

The Copiale Cipher - Clustering

Automatic Clustering of cipher letters

- For each letter **x** a vector is created that captures the distributions of letters that preceed **x**.
- If **x** is preceded 12 times by a, 2 times by b 0 times by c, 3 times by d → $[12, 2, 0, 3, ...]$
- Similarly, a vector is created for letters that follow **x**, and the two vectors are concatenated.
- Two letters are similar if the cosine distance $(1 - \cos\theta)$ between the corresponding vectors is small.



The Copiale Cipher- Decryption

- First Approach
  - Only Roman letters carry information
  - The resulting sequence does not make sense → substitution?
  - Automatic computer attacks → failed **but** gave slight preference for German

The Copiale Cipher- Decryption

- First Approach
  - Only Roman letters carry information
  - The resulting sequence does not make sense → substitution?
  - Automatic computer attacks → failed **but** gave slight preference for German
- Second Approach
  - Focus on German
  - Compare most common letters, digrams and trigrams
  - Use cluster map

The Copiale Cipher- Decryption

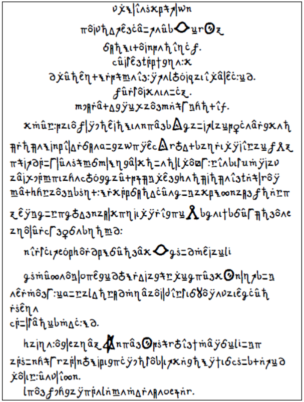
- First Approach
  - Only Roman letters carry information
  - The resulting sequence does not make sense → substitution?
  - Automatic computer attacks → failed **but** gave slight preference for German
- Second Approach
  - Focus on German
  - Compare most common letters, digrams and trigrams
  - Use cluster map
  - Use computer translator
  - Get help from native speakers

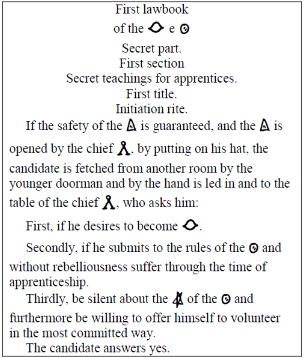
The Copiale Cipher- Decryption

- First Approach
  - Only Roman letters carry information
  - The resulting sequence does not make sense → substitution?
  - Automatic computer attacks → failed **but** gave slight preference for German
- Second Approach
  - Focus on German
  - Compare most common letters, digrams and trigrams
  - Use cluster map
  - Use computer translator
  - Get help from native speakers
  - Success!

The Copiale Cipher - Cracked!

- Describes the initiation into a secret society
- Some symbols still unknown





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.

Conclusion

- Traditional encryption schemes → not very safe
- Can squeeze out a lof of information from ciphertext
- Still challenging

Conclusion

- Traditional encryption schemes → not very safe
- Can squeeze out a lof of information from ciphertext
- Still challenging
- Still fun!