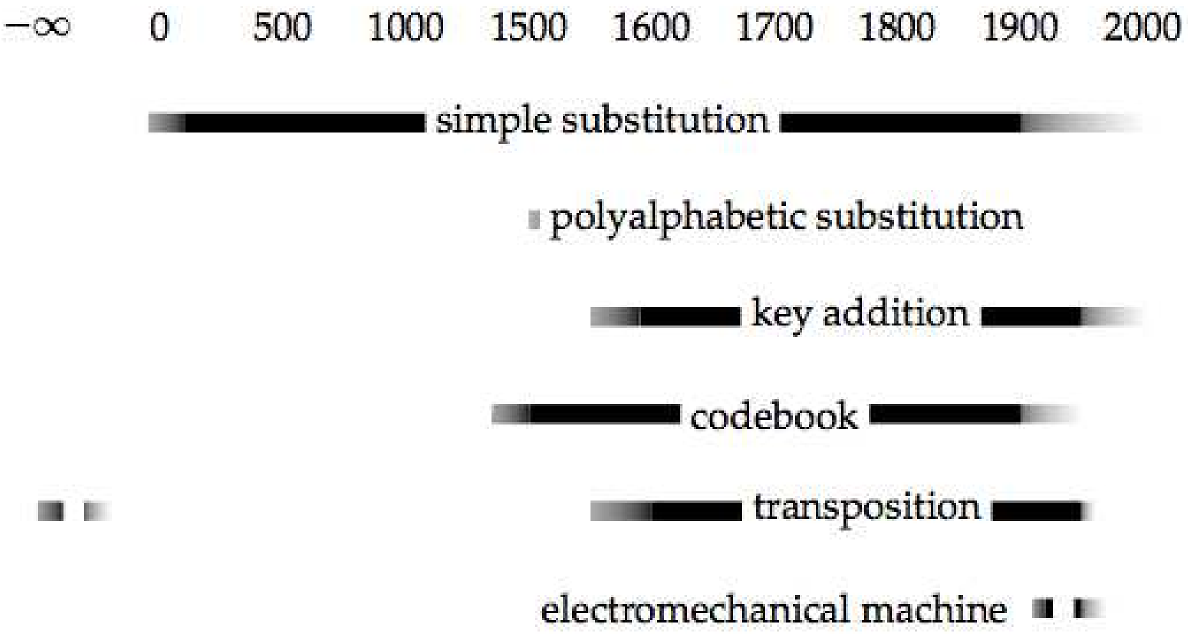
Desarrollo de la criptografía Con enfasis en Latino América

Universidad de la Népublica
Montevideo, Uruguay

Joachim von zur Gathen
Bonn, Alemania

Criptografía durante los milenios, antes computadores



Two fundamental cryptographic primitives:

- substitution,
- transposition.

Claude Shannon (1949): *confusion* and *diffusion*.
Goal: create enough of them to provide secrecy in communication.
Modern notions of cryptographic primitives: one-way and trapdoor functions.
Modern notions of security: 1970s.
Based on concepts from computational complexity.

Kerckhoffs' Principle

Auguste Kerckhoffs (1883) Principle:

> Il faut qu'il [le système] n'exige pas le secret, et qu'il puisse sans inconvénient tomber entre les mains de l'ennemi.
>
> It must not be required to keep the system secret, and it must be able to fall into the hands of the enemy without harm.

Still today, some companies believe in *security by secrecy* and develop systems in secrecy, which are in due course reverse-engineered and broken.

Famous example of failed *security by secrecy*: widely deployed *KeeLoq* system for protecting radio access to cars and garages. After being reverse-engineered, it was broken and the secret key can be recovered in seconds. The system is still in use!

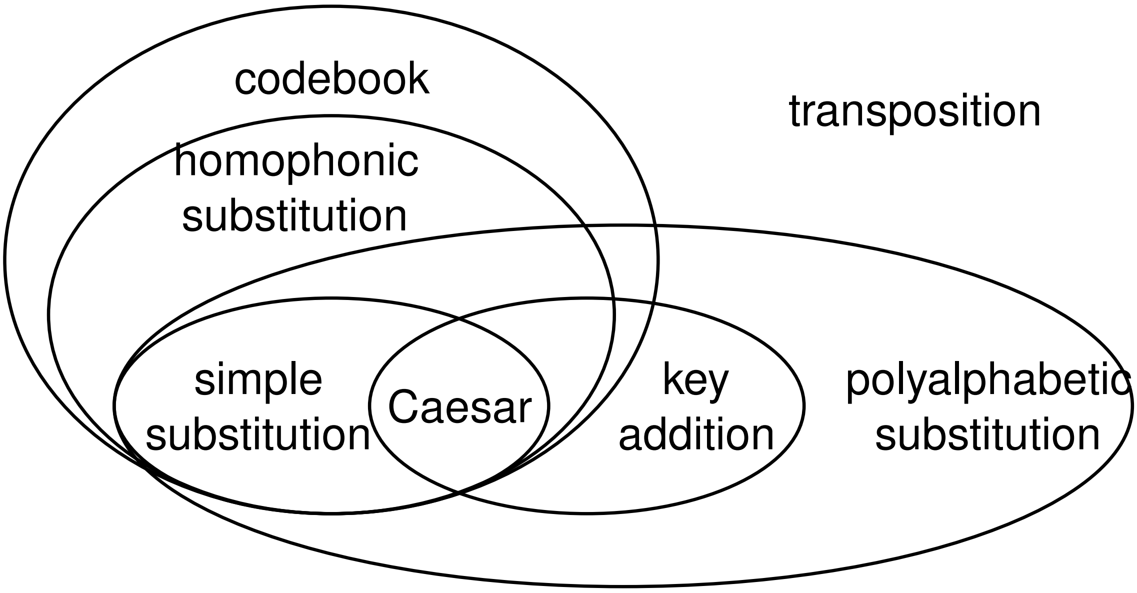


Figure : Taxonomy of classical cryptosystems.

Substitution: some *alphabet* X. E.g., the 26-letter English alphabet $A = \{a, b, c, \ldots, x, y, z\}$, or pairs of letters (*digrams*), so that $X = A^2$, or even longer *polygrams*, or bits $\mathbb{B} = \{0, 1\}$, or 128-bit words $X = \mathbb{B}^{128}$ for AES. In general, X is an arbitrary finite set. Furthermore, we have another alphabet Y, which might equal X or not.

- A *(simple) substitution* is a bijection $\sigma\colon A \longrightarrow Y$ from letters to some alphabet Y. Often $Y = A$.
- AES-128 uses two substitutions. The first is the fixed substitution SUBBYTES which consists of the patched inversion σ on $\mathbb{F}_{256}$ with $\sigma(x) = x^{-1}$ if $x \neq 0$ and $\sigma(0) = 0$, followed by a linear map, and is applied individually to each of the 16 blocks. The second one is the key addition $\sigma = \text{ADDKEY}\colon \{0,1\}^{128} \longrightarrow \{0,1\}^{128}$, where the 128-bit key (which we here consider as fixed) and the states are added bitwise.
- RSA with public key (N, e) is the substitution $\sigma\colon \mathbb{Z}_N \longrightarrow \mathbb{Z}_N$ with $\sigma(x) = x^e$.

A	B	C	D	E	F	G	H	I	J	K	L	M
0	1	2	3	4	5	6	7	8	9	10	11	12
N	O	P	Q	R	S	T	U	V	W	X	Y	Z
13	14	15	16	17	18	19	20	21	22	23	24	25

Table : Letter-to-number conversion in a 26-letter alphabet.

- The *Caesar cipher* identifies A with $\mathbb{Z}_{26}$ and uses the substitution $\sigma\colon \mathbb{Z}_{26} \longrightarrow \mathbb{Z}_{26}$ with $\sigma(x) = x + 3$. More generally, we might have any key $k \in \mathbb{Z}_{26}$ and use $\sigma_k(x) = x + k$. Caesar's Latin alphabet had 24 letters, identifying I = J and U = V. It is not clear from the historical accounts how the cyclic shift is handled. Is σ(Z) = C? This is what we would use in a modern sense. The later *Augustus cipher* σ_1, the shift by one letter, uses σ(Z) = AA. Today's modular arithmetic is younger by 1800 years, basically due to Gauß (1801).

Frequency analysis

Any simple substitution is easy prey to *frequency analysis*, if only the message is long enough. This cryptanalysis requires as its main tools frequency tables for individual letters, but also for *digrams* (pairs of letters), *trigrams* (triples), and short words. Below are are eight lists of letter frequencies in percent.

Frequency analysis

letter	COCA	HP	Ch.15	MM	D	F	E	I
a	8.28	7.94	7.83	8.04	5.26	7.75	12.25	10.71
b	1.54	1.59	1.53	1.54	1.85	0.99	1.48	0.74
c	3.18	1.99	2.91	3.06	3.62	2.67	3.63	5.14
d	3.9	4.96	3.47	3.99	5.05	3.35	5.33	3.73
e	12.17	11.91	12.27	12.51	17.41	16.61	14.01	12.04
f	2.14	2.05	2.70	2.30	1.50	1.08	0.46	1.29
g	2.12	2.57	1.68	1.96	2.94	1.29	1.05	1.82
h	5.03	6.73	4.43	5.49	5.90	0.93	1.22	1.81
i	7.29	6.21	8.80	7.26	8.85	7.33	5.50	10.26
j	0.19	0.11	0.10	0.16	0.15	0.71	0.64	
k	0.82	1.20	0.23	0.67	1.13			0.01
l	4.22	4.36	4.14	4.14	3.75	4.90	5.45	5.78
m	2.53	2.21	2.88	2.53	3.19	3.28	2.73	2.98
n	7.08	6.51	7.49	7.09	10.71	7.61	6.63	6.60
o	7.55	7.80	7.47	7.60	1.93	6.92	9.93	9.55
p	2.08	1.66	2.42	2.00	0.37	2.53	2.17	2.79
q	0.10	0.13	0.29	0.11	0.02	1.47	1.99	0.82
r	6.24	6.46	6.34	6.12	6.65	6.57	6.17	6.44
s	6.79	5.88	6.47	6.54	6.14	7.56	7.68	5.61
t	9.01	8.67	9.17	9.25	5.79	6.54	3.77	5.74
u	2.76	2.91	2.66	2.71	3.86	6.62	4.86	3.60
v	1.03	0.87	1.14	0.99	0.76	2.22	1.09	2.01
w	1.84	2.51	1.54	1.92	2.01			0.02
x	0.20	0.11	0.51	0.19	0.01	0.37	0.02	0.04
y	1.80	2.57	1.40	1.73	0.01	0.22	1.53	0.02
z	0.12	0.08	0.12	0.09	1.15	0.48	0.40	0.45
$\sum f_i^2$	6.50	6.36	6.65	6.58	7.77	7.55	7.51	7.22

- The *Vigenère cipher* with an l-letter keyword k uses l Caesar substitutions $\sigma_0, \ldots, \sigma_{l-1}$. Alternatively, it can be viewed as a simple substitution $\sigma\colon A^l \longrightarrow A^l$ with $\sigma(x) = x + k$, using letter-wise addition in A^l with truncation at the end. Example: keyword k = KEY of length $l = 3$, plaintext x = confuse the enemies yields encryption:

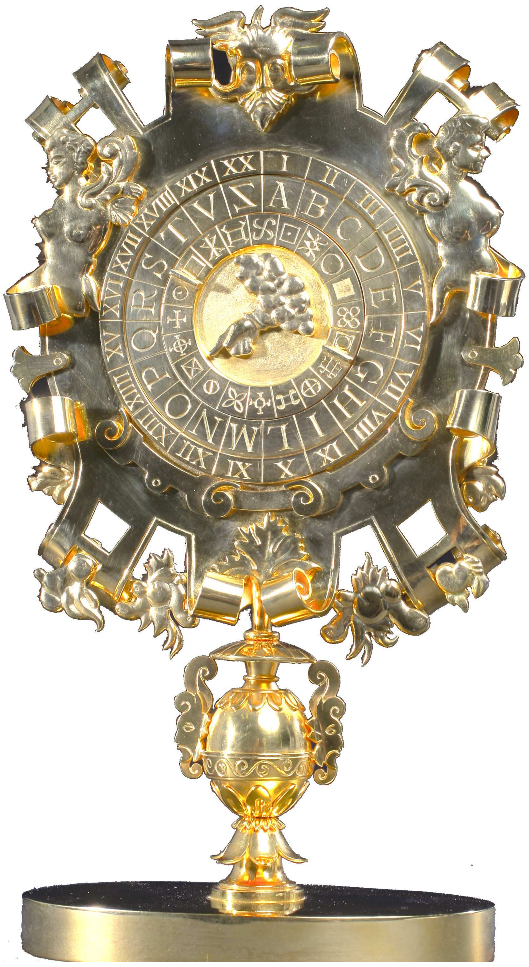
$$\begin{array}{rcl} x & = & \text{CONFUSETHEENEMIES} \\ k' & = & \text{KEYKEYKEYKEYKEYKE} \\ y & = & \text{MSLPYQOXFOILOQGOW} \end{array}$$

  As is common in classical cryptography, blanks, spaces, and punctuation marks are ignored. Thus $\sigma(x) = x + k'$, where k' is the 17-letter key obtained from the Vigenère key $k = key$ by sufficiently many repetitions with the *Procrustes rule* of cutting off unneeded key letters at the end.

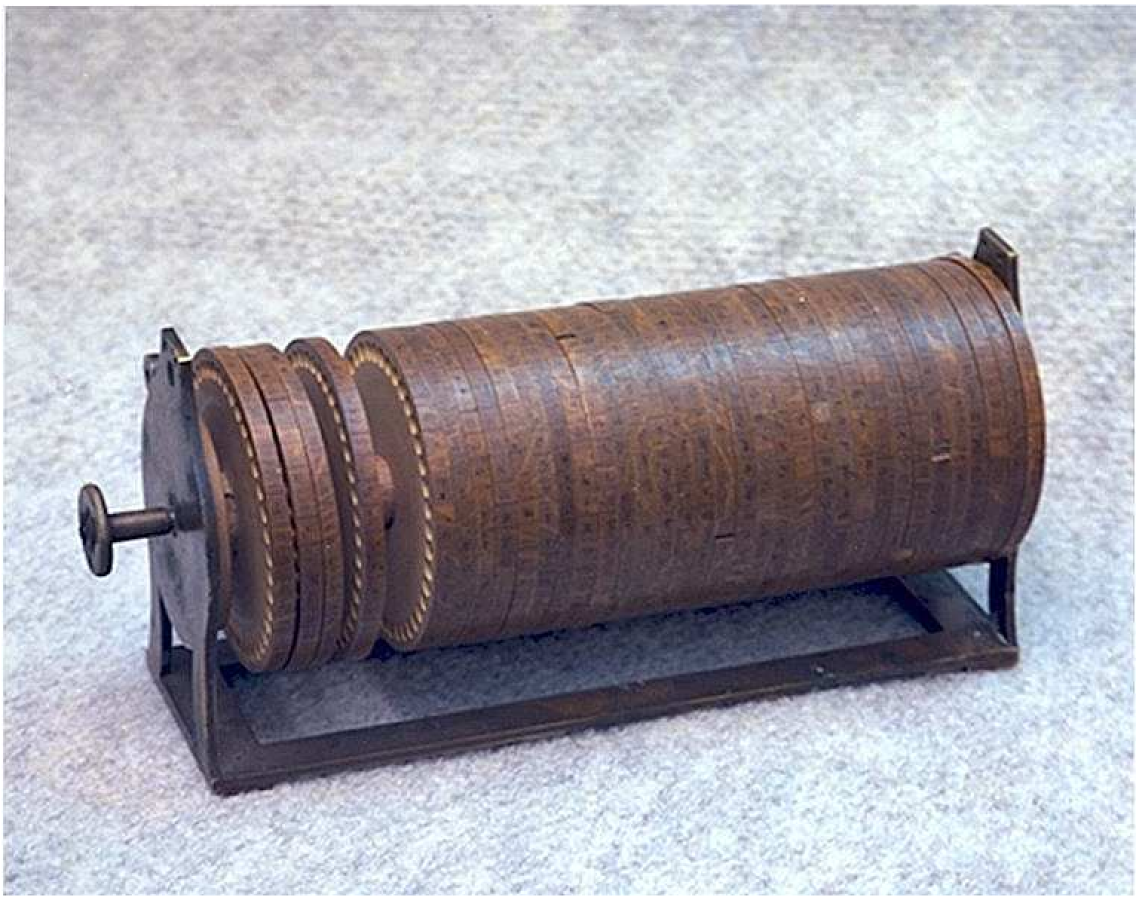
The attentive reader has noticed that in this short example, we have no fewer than four single-letter additions e + k = o. This is a general phenomenon, although usually not this frequent, and can be used to break this cryptosystem.

Giambattista della Porta (1535-1615): cipher disk

Mechanical implementations:

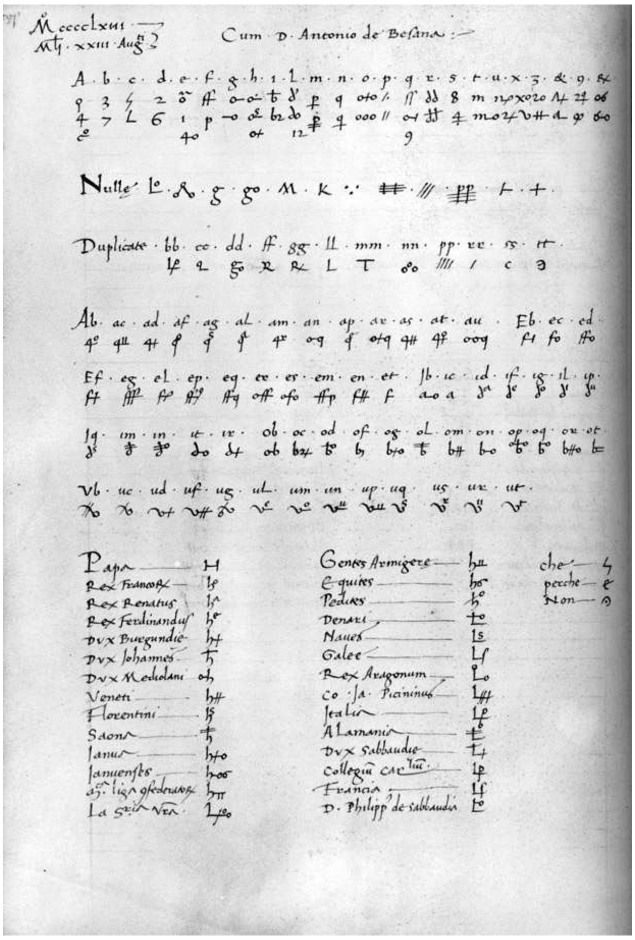


Thomas Jefferson: cipher wheel



- A *polyalphabetic substitution* applies a fixed sequence of simple substitutions one after the other. When the sequence is exhausted, one starts again with the first one.
- A *homophonic substitution* works in the same way, only for each letter we do not have just a single possibility, but several ones. We see an example in Tranchedino's codebook from 1463. Its first line (after the heading) gives the 21 letters A, b, . . ., z of the alphabet, plus the frequent words and, with, and of. Five of the letters get three possible encryptions, the others two. In general, the goal of the multiple possibilities is to even out the disparate frequencies of the various letters. The corresponding σ is now only a relation, not necessarily a function.
  The Spanish cipher from around 1590 for Moreo contains the line: *Las nullas tendran una raya enzima, exemplo* $\overline{19}$. This provides a systematic way of introducing a large number of dummies.

Francesco Tranchedino (1411–1496): Nomenclator



M cccclxiiii
Mtl · xxiii · Aug

Cum · D · Antonio de Besana :

A · b · c · d · e · f · g · h · i · l · m · n · o · p · q · r · s · t · u · x · z · & · 9 ·

Nulle

Duplicate · bb · cc · dd · ff · gg · ll · mm · nn · pp · rr · ss · tt ·

Ab · ac · ad · af · ag · al · am · an · ap · ar · as · at · au · Eb · ec · ed ·

Ef · eg · el · ep · eq · er · es · em · en · et · Jb · ic · id · if · ig · il · ip ·

Jq · im · in · it · ix · ob · oc · od · of · og · ol · om · on · op · oq · or · ot

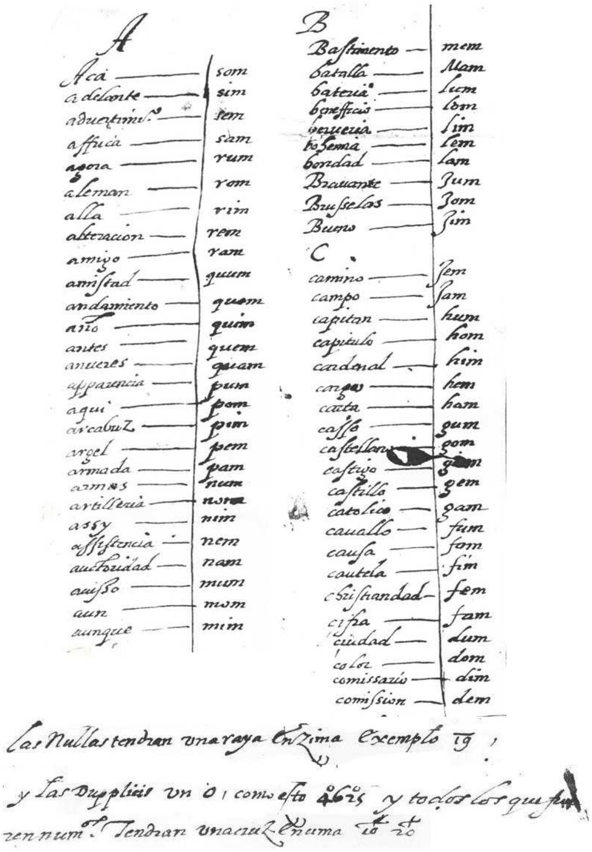
Vb · uc · ud · uf · ug · ul · um · un · up · uq · us · ur · ut ·

Papa
Rex Francorum
Rex Renatus
Rex Ferdinandus
Dux Burgundie
Dux Johannes
Dux Mediolani
Veneti
Florentini
Saona
Janua
Januenses
a3̄ · liga confederatorum
La S^{ria} Vra ·

Gentes Armigere
Equites
Pedites
Denari
Naues
Galee
Rex Aragonum
Co · Ja · Picininus
Jtalia
Alamania
Dux Sabbaudie
Collegium car^{lium}
Francia
D · Philipp de sabbaudia

che
perche
Non

Spanish codebook for Moreo, 16th century



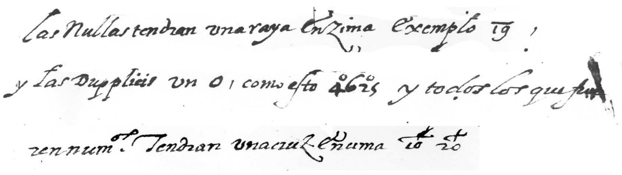
A	
Aca	som
adelante	sim
aduertimi.^to^	sem
affrica	sam
agora	rum
aleman	rom
alla	rim
alteracion	rem
amigo	ram
amiſtad	quum
andamiento	quom
año	quim
antes	quem
anueres	quam
apparencia	pum
aqui	pom
arcabuz	pim
argel	pem
armada	pam
armas	num
artilleria	nom
assy	nim
aſſiſtencia	nem
auctoridad	nam
auiſo	mum
aun	mom
aunque	mim

B	
Baſtimento	mem
batalla	llam
bateria	lum
beneffiçio	lom
berueria	lim
bohemia	lem
bondad	lam
Brauante	Jum
Bruſelas	Jom
Bueno	Jim
C	
camino	Jem
campo	Jam
capitan	hum
capitulo	hom
cardenal	him
cargos	hem
carta	ham
caſſo	gum
caſtellano	gom
caſtigo	gim
caſtillo	gem
catolico	gam
cauallo	fum
cauſa	fom
cautela	fim
christiandad	fem
cifra	fam
ciudad	dum
color	dom
comissario	dim
comiſsion	dem

las Nullas tendran una raya Enzima Exemplo 19 ,

y las Dupplicis vn o, como esto 4625 y todos los que fue-
ren num.^os^ Tendran una cruz Encima 10 20

Spanish codebook for Moreo, 16th century

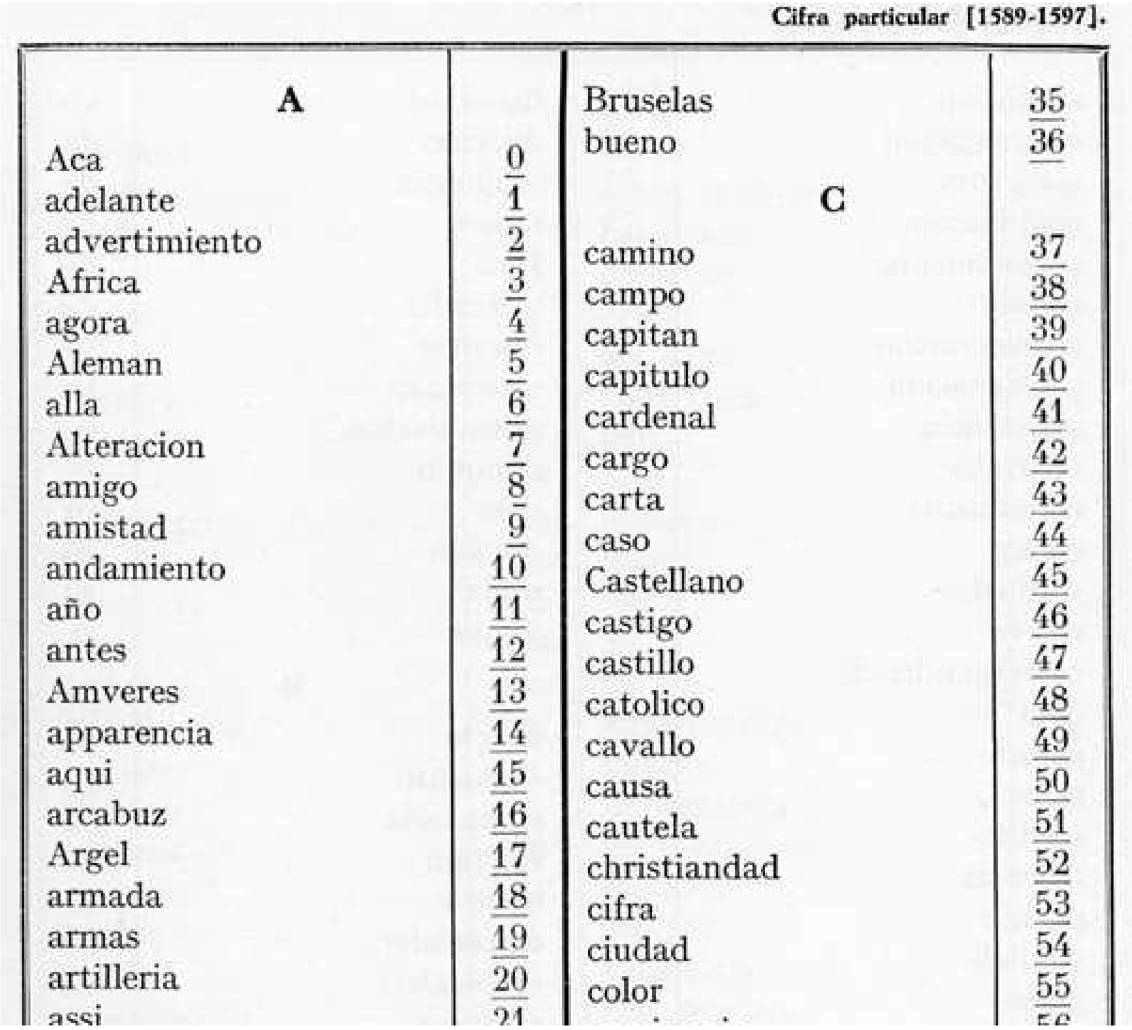


las Nullas tendran una raya enzima exemplo $\overline{19}$;

y las Dupplicis un o, como esto $\overset{o}{4}6\overset{o}{2}5$ y todos los que fue-

ren num.os Tendran una cruz enzima $\overset{+}{10}$ $\overset{+}{20}$

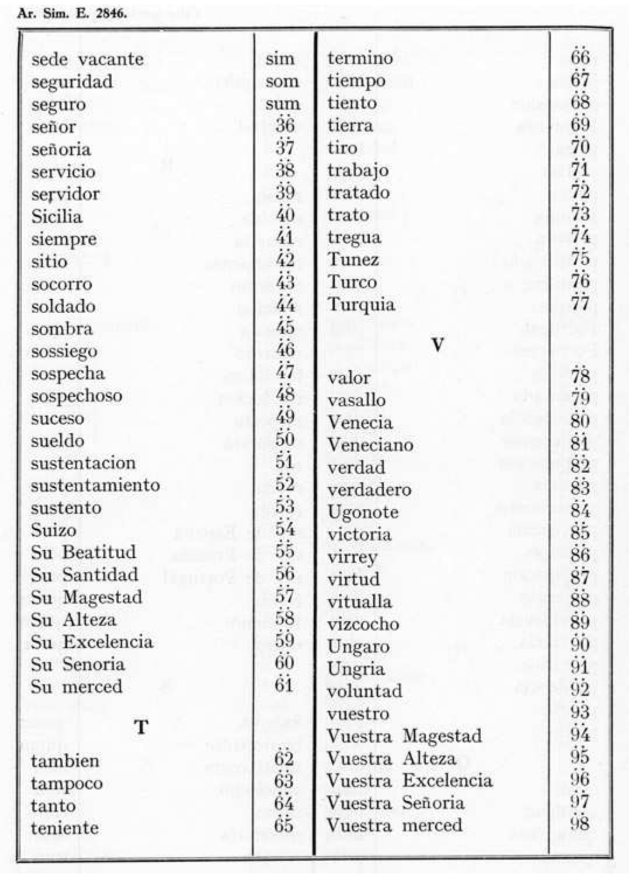
Spanish codebook 16th century



Cifra particular [1589-1597].

A		Bruselas	35
Aca	0	bueno	36
adelante	1	**C**	
advertimiento	2	camino	37
Africa	3	campo	38
agora	4	capitan	39
Aleman	5	capitulo	40
alla	6	cardenal	41
Alteracion	7	cargo	42
amigo	8	carta	43
amistad	9	caso	44
andamiento	10	Castellano	45
año	11	castigo	46
antes	12	castillo	47
Amveres	13	catolico	48
apparencia	14	cavallo	49
aqui	15	causa	50
arcabuz	16	cautela	51
Argel	17	christiandad	52
armada	18	cifra	53
armas	19	ciudad	54
artilleria	20	color	55
assi	21		56

Spanish codebook 16th century



Ar. Sim. E. 2846.

sede vacante	sim	termino	6̇6̇
seguridad	som	tiempo	6̇7̇
seguro	sum	tiento	6̇8̇
señor	3̇6̇	tierra	6̇9̇
señoria	3̇7̇	tiro	7̇0̇
servicio	3̇8̇	trabajo	7̇1̇
servidor	3̇9̇	tratado	7̇2̇
Sicilia	4̇0̇	trato	7̇3̇
siempre	4̇1̇	tregua	7̇4̇
sitio	4̇2̇	Tunez	7̇5̇
socorro	4̇3̇	Turco	7̇6̇
soldado	4̇4̇	Turquia	7̇7̇
sombra	4̇5̇		
sossiego	4̇6̇	V	
sospecha	4̇7̇	valor	7̇8̇
sospechoso	4̇8̇	vasallo	7̇9̇
suceso	4̇9̇	Venecia	8̇0̇
sueldo	5̇0̇	Veneciano	8̇1̇
sustentacion	5̇1̇	verdad	8̇2̇
sustentamiento	5̇2̇	verdadero	8̇3̇
sustento	5̇3̇	Ugonote	8̇4̇
Suizo	5̇4̇	victoria	8̇5̇
Su Beatitud	5̇5̇	virrey	8̇6̇
Su Santidad	5̇6̇	virtud	8̇7̇
Su Magestad	5̇7̇	vitualla	8̇8̇
Su Alteza	5̇8̇	vizcocho	8̇9̇
Su Excelencia	5̇9̇	Ungaro	9̇0̇
Su Senoria	6̇0̇	Ungria	9̇1̇
Su merced	6̇1̇	voluntad	9̇2̇
		vuestro	9̇3̇
T		Vuestra Magestad	9̇4̇
tambien	6̇2̇	Vuestra Alteza	9̇5̇
tampoco	6̇3̇	Vuestra Excelencia	9̇6̇
tanto	6̇4̇	Vuestra Señoria	9̇7̇
teniente	6̇5̇	Vuestra merced	9̇8̇

- *Nomenclators* or *codebooks* have large alphabets X and Y, with several hundreds (in the 17th century) or thousands (19th century) of elements each. Y has at least as many elements as X does, and the codebook is a simple substitution $\sigma\colon X \longrightarrow Y$, or a relation if homophones are used, as is often the case. The alphabet X usually comprises letters plus certain frequently occurring items, such as syllables, words, or names that were likely to appear in the correspondence. Their use is recorded from 1377 to the Second World War, where in one German submarine cipher each square of a grid covering the North Atlantic was given its code.

- The basic ingredient of the *one-time pad* is a substitution σ on single bits, where a (random) key bit k is chosen in $\mathbb{B}$, and a one-bit message $x \in \mathbb{B}$ is encrypted as $\sigma(x) = x + k$. Longer messages are encrypted by repeating this procedure, with keys chosen anew (independently at random) for each message bit. It is unconditionally secure, but highly impractical. The *two-time pad* is insecure; *VENONA*.
- The *Playfair* cipher is a simple substitution $\sigma\colon A_0^2 \longrightarrow A_0^2$ on digrams, where $A_0 = A \setminus \{\text{j}\}$ is the standard alphabet with the letter j removed.
- In the *Enigma*, the secret key determines (in a complicated fashion) a sequence of simple substitutions $\sigma_0, \sigma_1, \ldots,$ with $\sigma_i\colon A \longrightarrow A$ for all i.

Latin American cryptography: Cortés

Hernán Cortés (1485–1547), the Spanish *conquistador* of Aztec Mexico, wrote to his representative Francisco Nuñez in Spain on 25 June 1532 from Cuernavaca in Mexico. This letter deals with Cortés' financial arrangements—real estate deals involving whole provinces. Most of the text on its last four pages is encrypted, and was presumably decrypted at the time. However, the *Archivo de Indias* at Sevilla conserves only the original, which resisted several attempts at deciphering. It may well be the first documented cryptography in America.

The *Museo Nacional* in Mexico published in 1924 in the leading local newspapers a competition to decipher it—an early forerunner of today's crypto challenges. Francisco Monterde García Icazbalceta succeeded, and was presented the prize of 200 pesos in November 1925. The cipher had $1, 2$, or 3 symbols (49 in total) for each of the 24 letters, plus *ll* and *que*. Monterde solved it by frequency analysis, aided by the separation of words. He explains how difficult the first inroads were, and how easy the rest:

El resto fue sencillo: pude leer la carta con facilidad, reemplazando todos los signos por letras; pero había tenido que emplear varias horas de la noche, durante tres meses, para llegar a ese resultado.

Latin American cryptography: Cortés

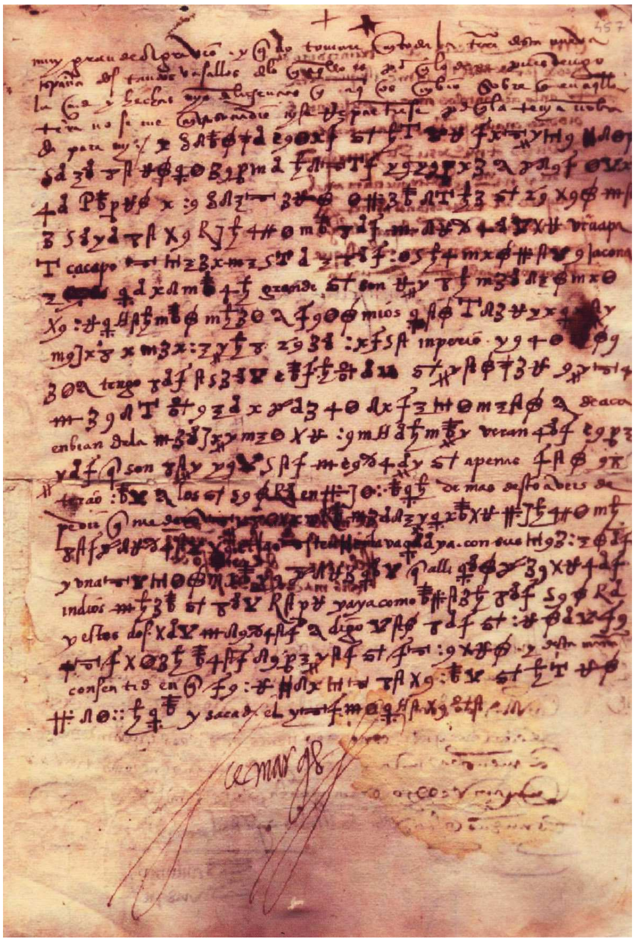


Figure : The fourth page of Hernán Cortés' letter from 25 June 1532

Latin American cryptography: Cortés

Shown is the last of Cortés' four pages. The first lines of the ciphertext, beginning on line 5, read:

[...] Y quando veais q ay se siente quán-
to yo lo encarezco, aueis de dezir q pues así
lo hazen y me quiren agrauiar q me den po-
r todo lo de Guaxaca los pueblos de Vruapa
y Cacapo e Tiripitío y los Matalcingos e Jacona
y [word crossed out] Coyuca la grande q son en la c [*p*] rouincia
de Mechoacán c [*p*]ara q sean mios con yuredic[i]on
ceuil y criminal mero misto inperio de la [ge]ne-
ral tengo los otros vasollos q nonbré en el
preuillejo y por la uisitación q de aca
enbian dela prouincia de Mechoacán veran los vezi-
nos q son [son] destos pveblos q apenas son el
tercio más q los q tengo en Guaxaca. [...]

Latin American cryptography: Martí

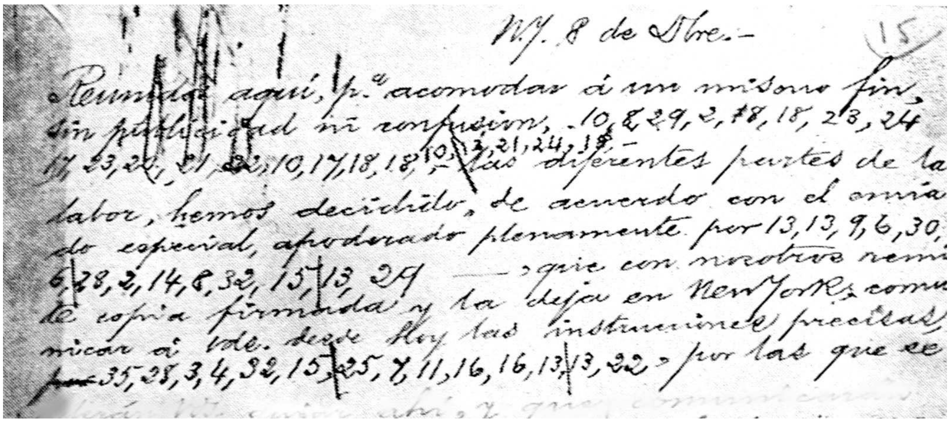
José Julián Martí Pérez (1853–1895) was the Cuban national hero who led the uprising against Spanish colonial occupation in February 1895. He was consul of Uruguay in New York 1879–1892.

Latin American cryptography: Martí

In a letter to his fellow conspirators in Cuba, he used a Vigenère cipher with key word habana $= 8, 1, 2, 1, 15, 1$ and his 28-letter alphabet a $= 1$, b, $\ldots$, l, ll, m, n, ñ, o, $\ldots$. We show the first lines of his famous *Plan de Alzamiento para Cuba*, written on 8 December 1894 in his New York City exile with two fellow revolutionaries:

N.Y. 8 de Dbre. Reunidos aquí, para acomodar á un mismo fin, sin publicidad ni confusión, *b.g.ya con un término cierto* —las diferentes partes de la labor, hemos decidido, de acuerdo con el enviado especial, apoderado plenamente por *el general Gomez*–, que con nosotros remite copia firmada y la deja en New York, comunicar á Vds. desde hoy las instrucciones precisas, *y ya como finales*, por las que se deberán Vds. guiar ahí.

Latin American cryptography: Martí



N.Y. 8 de Dbre.— 15

Reunidos aquí, p.a acomodar á un mismo fin, sin publicidad ni confusion, 10,8,29,2,18,18,23,24 17,23,20,21,22,10,17,18,18, 10,13,21,24,13 — las diferentes partes de la labor, hemos decidido, de acuerdo con el enviado especial, apoderado plenamente por 13,13,9,6,30,6,28,2,14,8,32,15,13,29 ———, que con nosotros remite copia firmada y la deja en New York, comunicar á Vds. desde hoy las instrucciones precisas, ~~por~~ 35,28,3,4,32,15,25,7,11,16,16,13,13,22, por las que se

Figure : The first lines of Martí's plan for the Cuban uprising

Latin American cryptography: Martí

Martí commits a cardinal sin in cryptography: mixing plaintext and ciphertext. But not even such a blunder could stop Cuban independence. The cleartext was written first, leaving space for the ciphertext which was then inserted by another hand. On line 3, too little, and on line 6, too much space had been left within the plaintext.

Periods

antiquity	1500 BC – 100 AD
Arab civilization	800 – 1400
European Middle Ages	1000 – 1500
Renaissance	1450 – 1600
Baroque, salon cryptography	1600 – 1850
mechanical devices	1580 – 1950
electromechanical devices	1920 – 1950
computers	1943 – present
public key systems	1976 – present

Table : Cryptographic time periods.

For the second cryptographic primitive, we have a length parameter l. A *transposition* is simply a bijection (or permutation) on the first l numbers:

$$\tau\colon \{0, \ldots, l-1\} \longrightarrow \{0, \ldots, l-1\}.$$

- AES uses the transposition SHIFTROWS which performs certain cyclic shifts on the rows of the state matrix.

- In a single columnar transposition we write the plaintext in r rows of length c and read it off in columns as the ciphertext. Thus $x =$ COLUMN becomes $y =$ CLMOUN $= x_0x_2x_4x_1x_3x_5$ in an $r \times c = 3 \times 2$ array:

$$x = \begin{array}{cc} \text{C} & \text{O} \\ \text{L} & \text{U} \\ \text{M} & \text{N} \end{array}$$

The transposition τ and its inverse α are given by

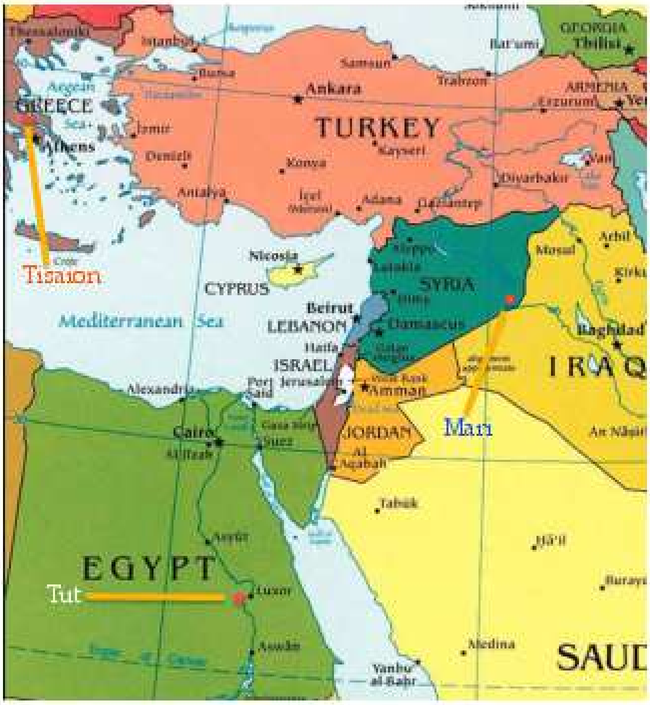
i	0	1	2	3	4	5
$\tau(i)$	0	3	1	4	2	5
$\alpha(i)$	0	2	4	1	3	5

and one checks that $\tau(i) = 3i - 5\lfloor i/2 \rfloor$.

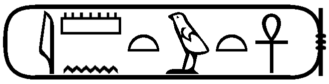
- The Spartan *skytale* is a columnar transposition.

From a transposition τ on l numbers we obtain, for any alphabet A, a substitution $\tau_A \colon A^l \longrightarrow A^l$ by setting $\tau_A(x) = x_{\alpha(0)} x_{\alpha(1)} \cdots x_{\alpha(l-1)}$ for $x \in A^l$, where α is the inverse of τ. Thus a transposition of length l yields a simple substitution on l-grams. However, it is profitable to keep the two primitives apart. For one, τ as above is much less “powerful” than a general substitution on A^l, and furthermore, τ works for any A and might be called a “scheme” for such substitutions. From a higher point of view, substitutions are semantic objects and transpositions have a syntactical (or combinatorial) nature.

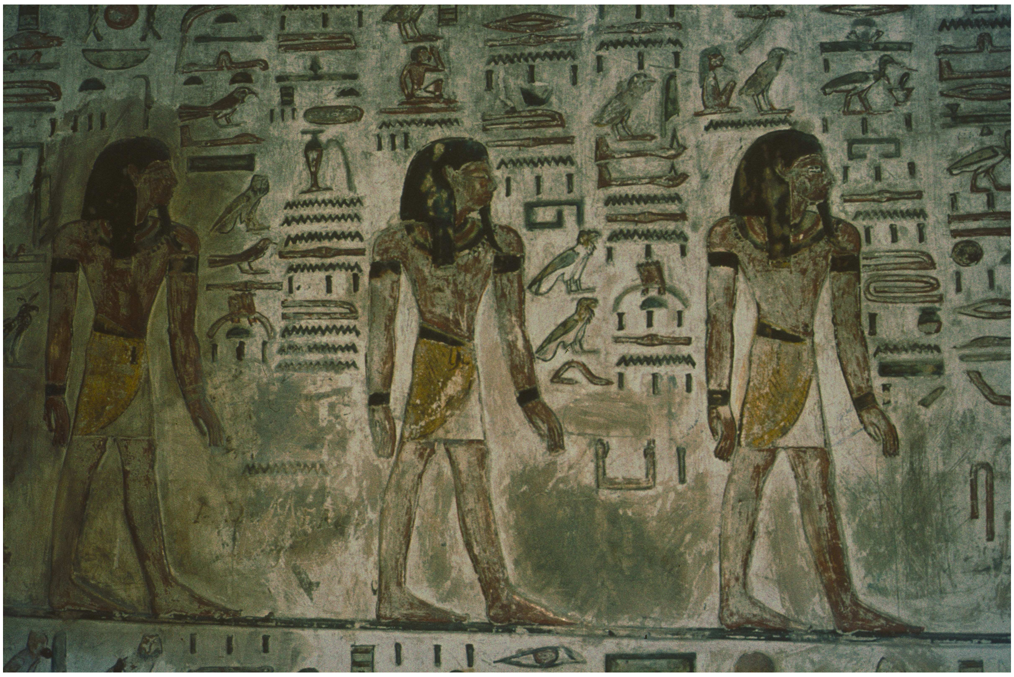
The classical world: Near East and Europe



Tutankhamun, c. 1343–1323 BC

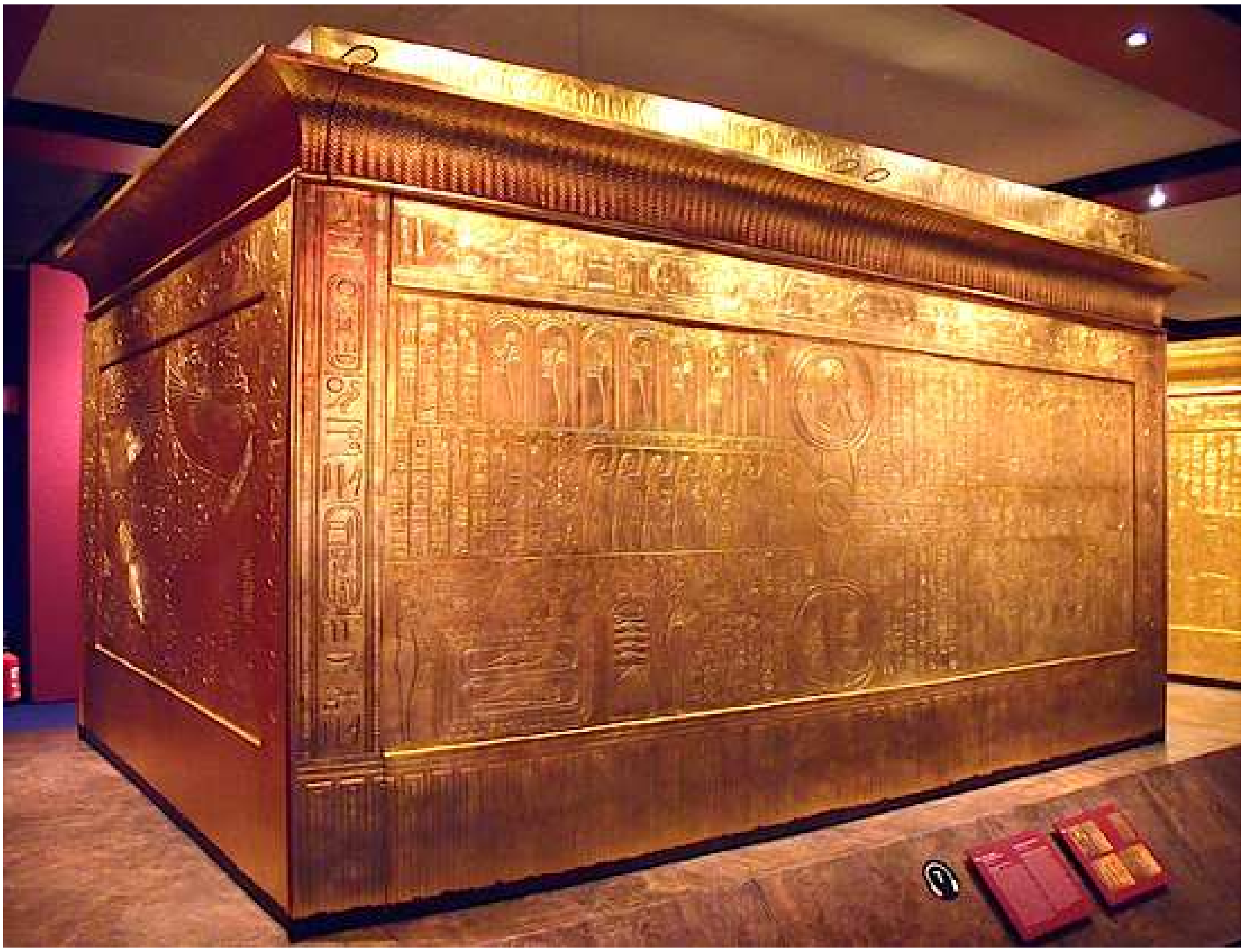


Tutankhamun’s tomb: one wall

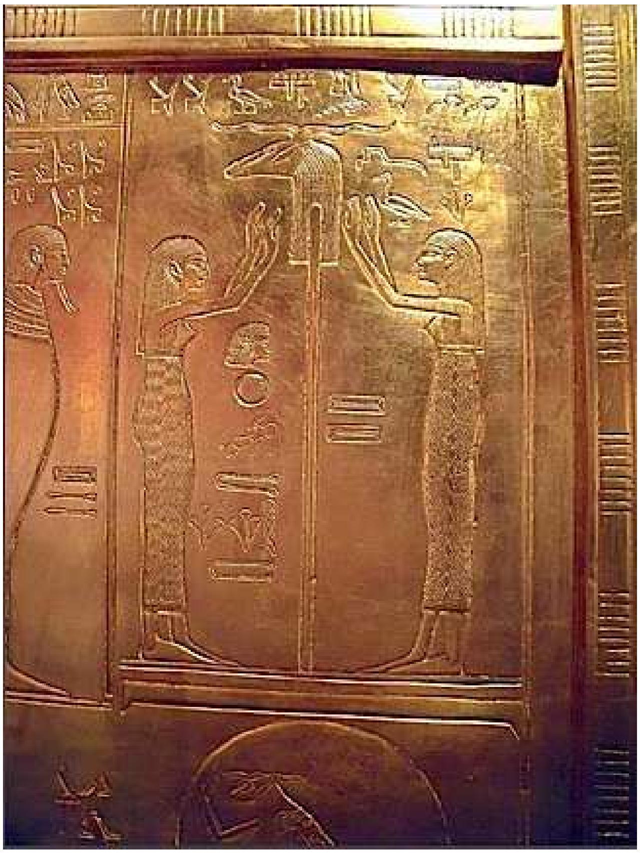


The usual hieroglyphic writing system of Ancient Egypt employed symbols at three levels: meaning, word, and sound. As an example, the red crown of lower Egypt 𓋔 can mean this crown, or be placed after other hieroglyphs spelling this crown letter by letter to categorize them. It is pronounced *dšrt* or *nt*, and can also (rarely) stand for the letter *n*. It is a remarkable achievement—akin to the cryptanalysis of an unknown system in an unknown language—that starting with Jean-François Champollion (1790–1832), egyptologists have learned how to understand and pronounce hieroglyphic texts.

Tutankhamun’s second shrine

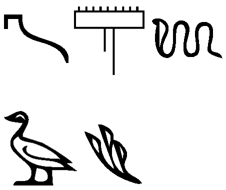


Tutankhamun’s second shrine.
Detail at top right: Isis, Nephtys, Ra



Tutankhamun’s second shrine.
Detail at top right: Isis, Nephtys, Ra

Inscription in lines 2 and 3:



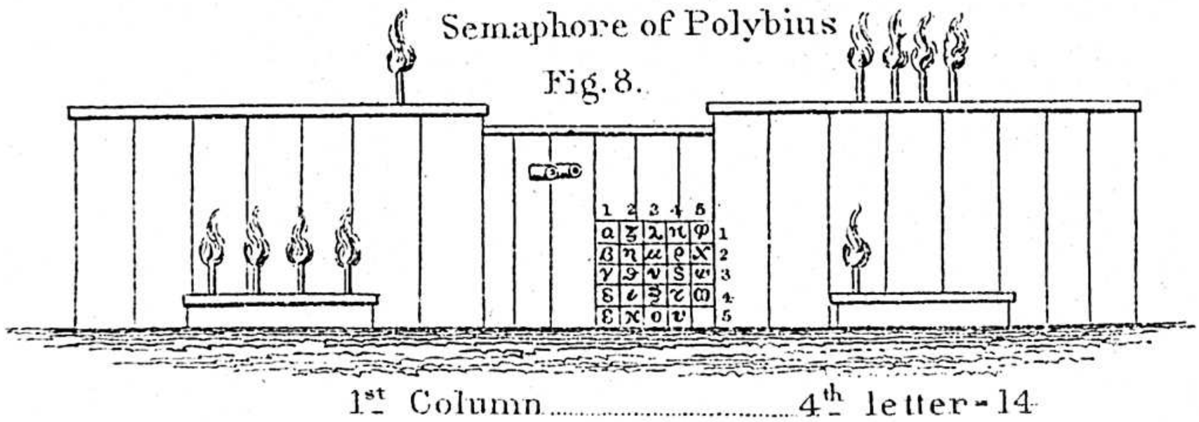
Literally: n ms hfrw sa hn
Decipherment: sn m-ha-f = these are around him [for protection]
Substitution and transposition.
Claude Shannon 1949: confusion and diffusion.

Euphrates near Mari, today in Syria

Mari: watch towers, c. 2500 BC

Mari

Polybios’ signalling device



Semaphore of Polybius

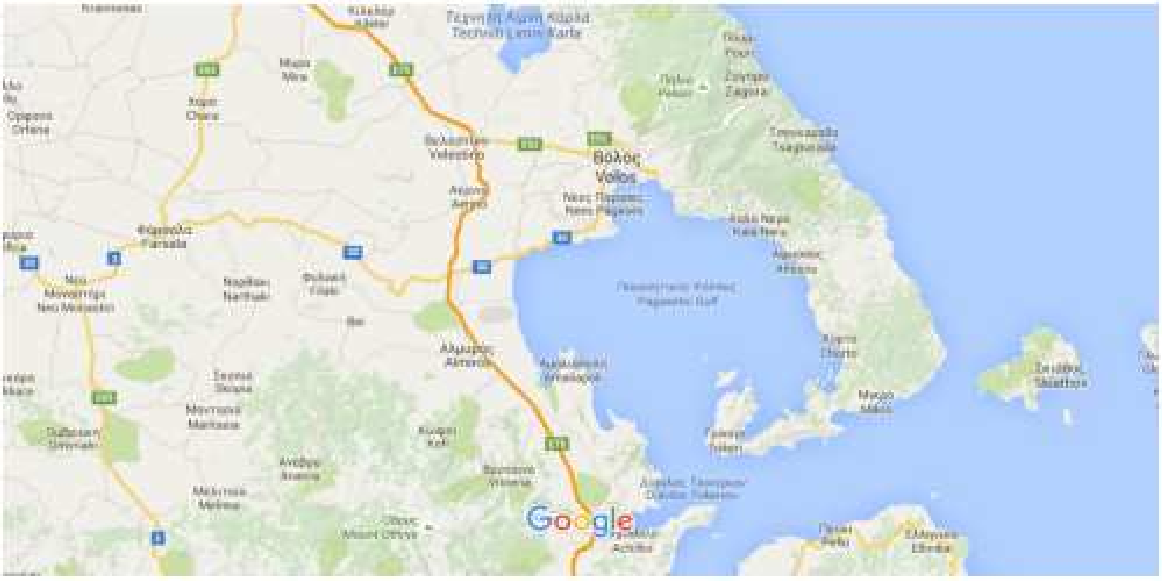
Fig. 8.

	1	2	3	4	5
1	α	ζ	λ	π	φ
2	β	η	μ	ρ	χ
3	γ	θ	ν	ς	ψ
4	δ	ι	ξ	τ	ω
5	ε	κ	ο	υ	

1st Column.......................................4th letter - 14

The bay of Tisaion

7 km over water to Demetrios, 140 km over land.
King Philip V. of Macedonia, 238–179 BC.
Second Macedonian war.


Τεχνητή Λίμνη Κάρλα
Technití Límni Kárla
Βελεστίνο
Velestino
Βόλος
Volos
Φάρσαλα
Farsala
Παγασητικός Κόλπος
Pagasetic Gulf
Αλμυρός
Almyros
Σκιάθος
Skiathos
Google