Compression, Correction, Confidentiality, and Comprehension

Steven M. Bellovin
smb@cs.columbia.edu
http://www.cs.columbia.edu/~smb

+1 212-939-7149
Department of Computer Science
Columbia University

cc BY NC

Early Telegraphy

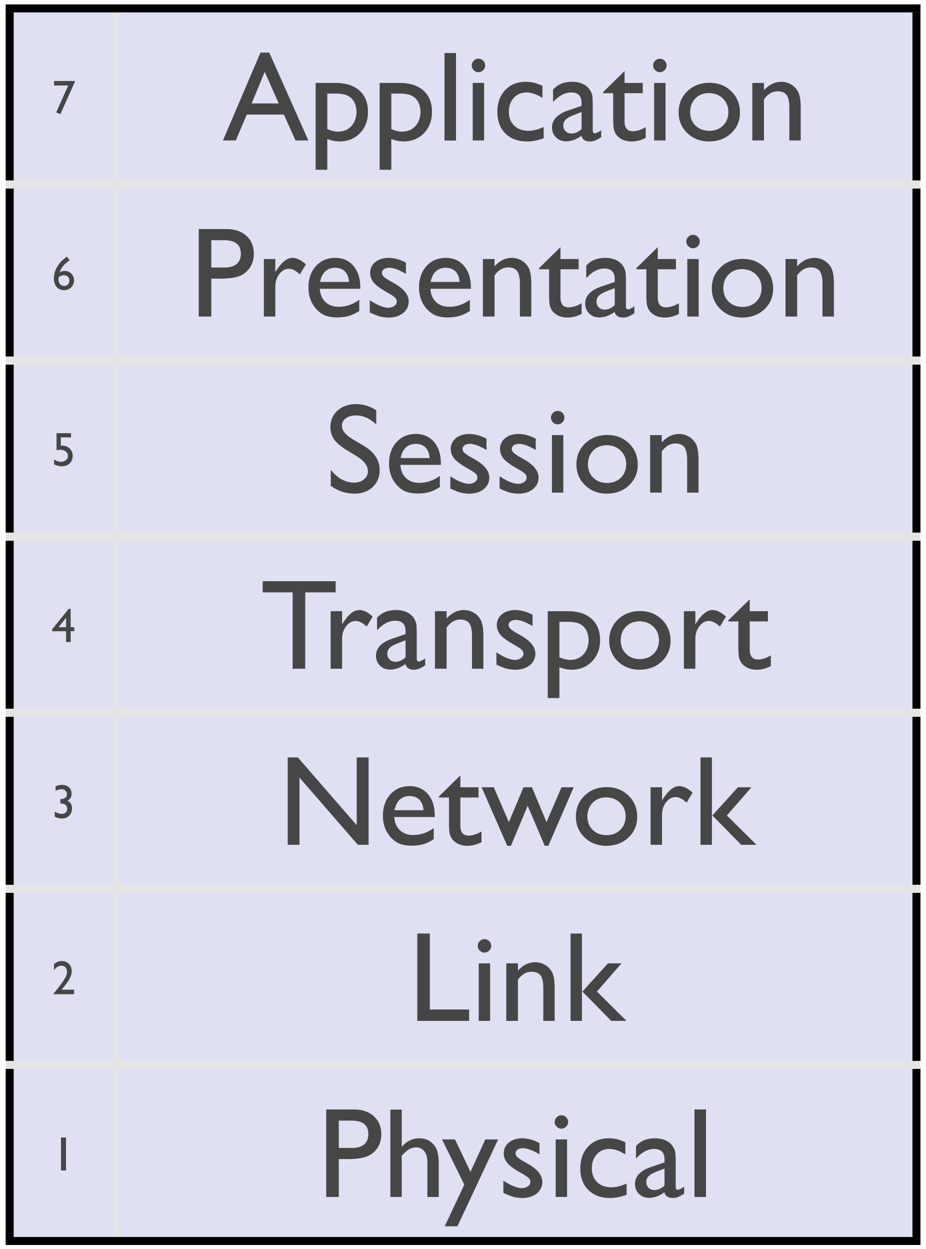
- Early telegraphy, especially overseas, was *very* expensive: $100 for twenty words trans-Atlantic in 1866.
- Messages were no longer sealed; a telegraph operator saw them
- The solution was *code books*
- Precedents: optical semaphore networks; naval signaling flags

Alfred Harrell, ©1974. Smithsonian Institution, http://www.si.edu
Image 74-2491. Used by permission.

Questions

- What did code books *really* do?
- How did they fit into the overall communications picture?
- Is there a relationship to modern network technology?

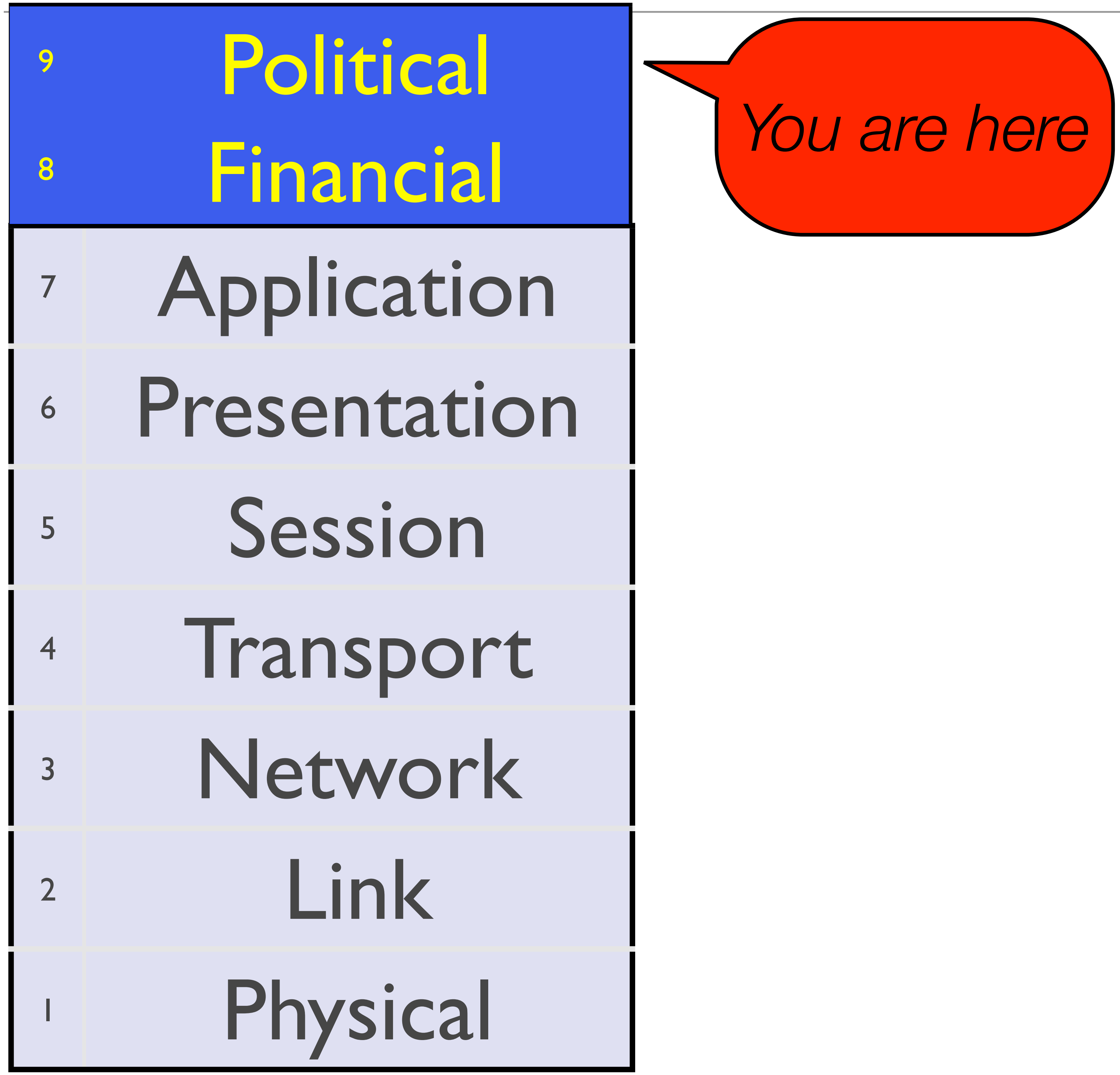
The Network Stack



7	Application
6	Presentation
5	Session
4	Transport
3	Network
2	Link
1	Physical

- Standardized model; doesn’t completely match the Internet
- Each layer provides services to the layer above
- Layers rely on the properties of the layer below
- Layers communicate with their peers on other nodes

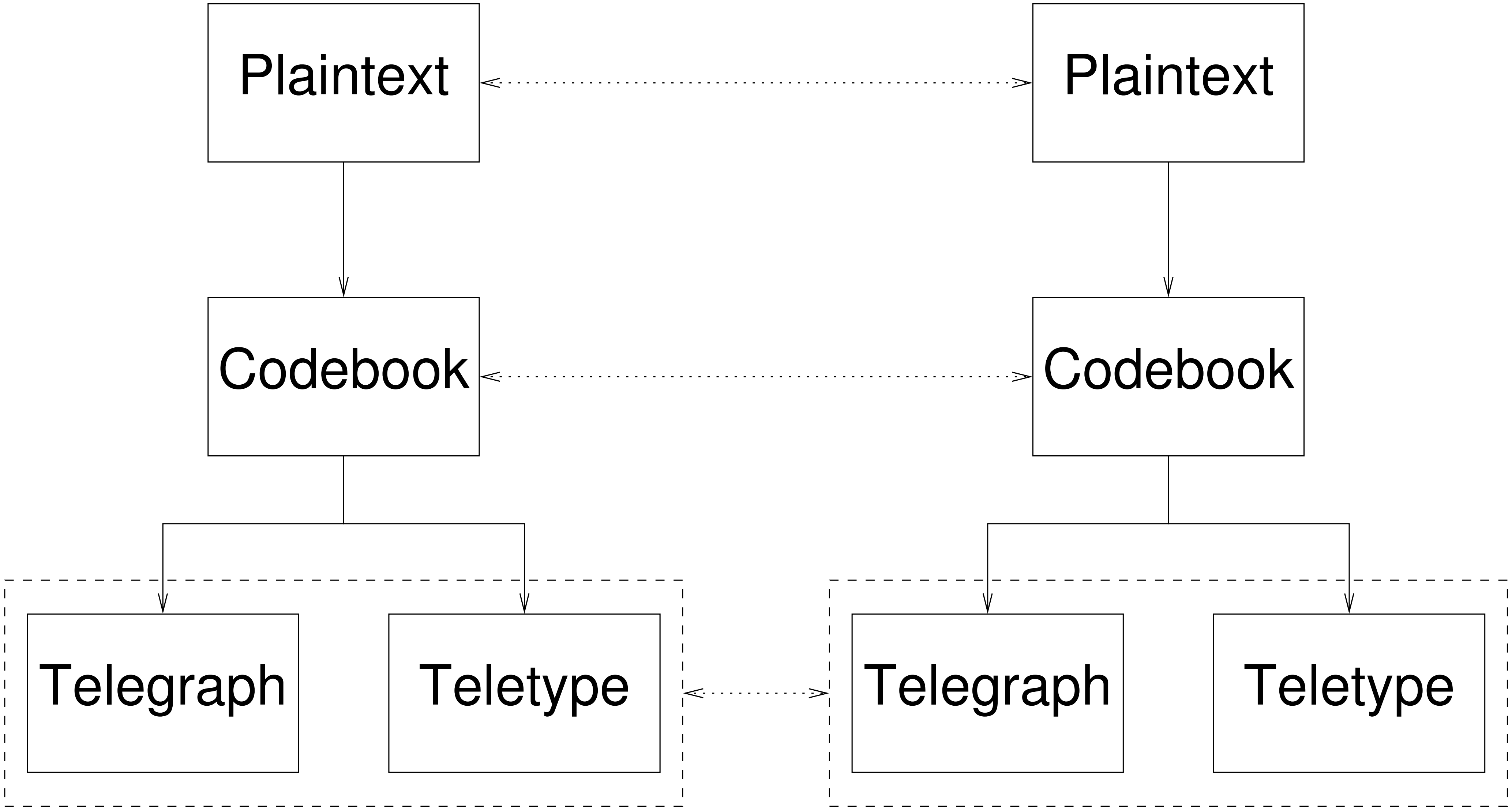
The *Real* Network Stack



9	Political
8	Financial
7	Application
6	Presentation
5	Session
4	Transport
3	Network
2	Link
1	Physical

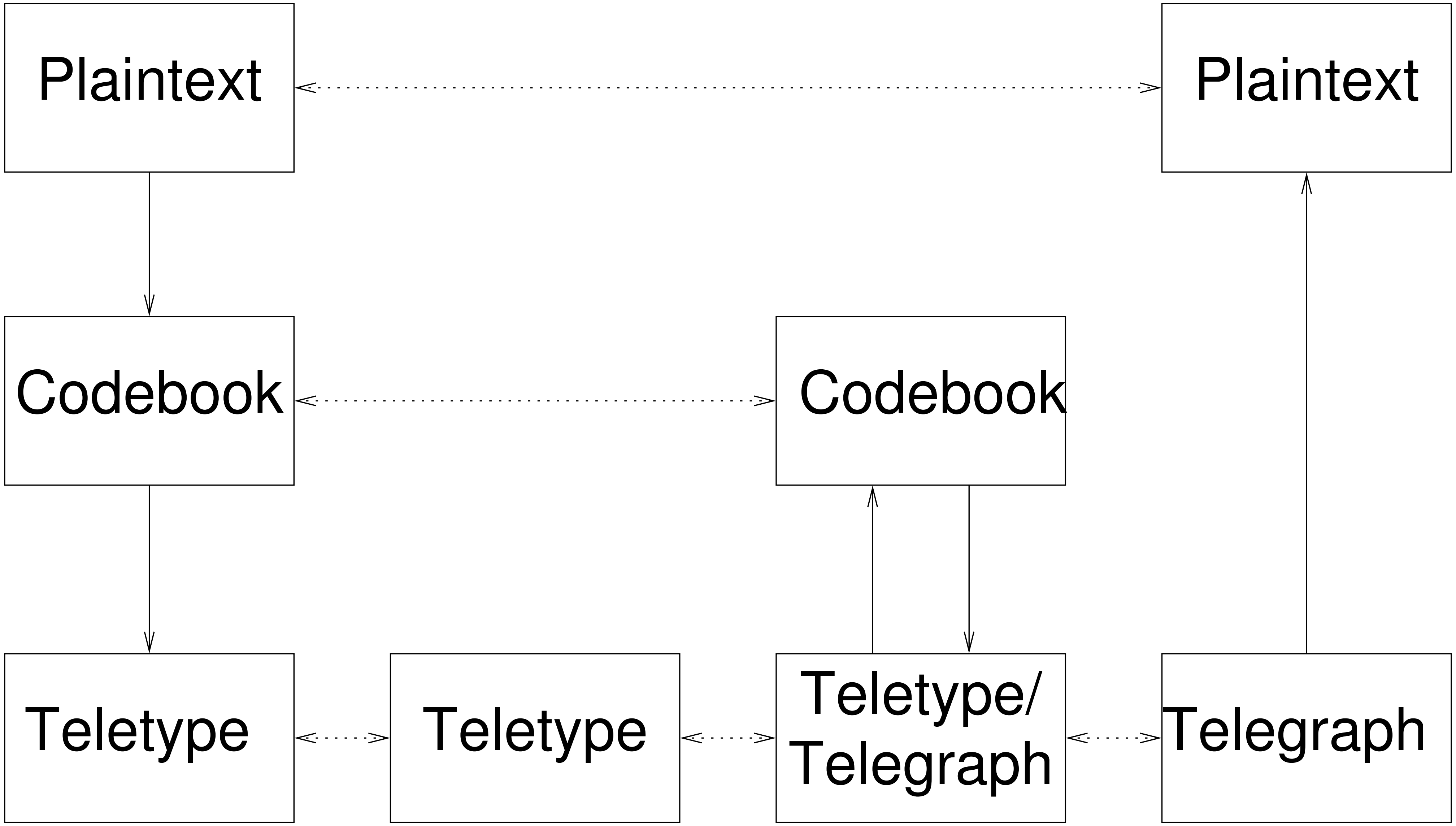
You are here

The Telegraph Stack



Note: link/network layers merged here; there could be many transmissions over different telegraph links

Decoding During Transmission



Some companies offered in-net decoding, to avoid problems from code book mismatches

Fitting Four Focus Areas to the Stack

- Compression — reducing transmission cost
- Correction — detecting and correcting errors
- Confidentiality — protecting the content of a message
- Comprehension — understanding other cultures, distant in time and space

Compression

Compression Metrics

- The goal was not to minimize characters sent, it was to minimize *cost*
  - ➡A layer 8 consideration
- Cost was affected by telegraph company tariffs and international regulations
  - ➡Layer 9?
- Permissible “words” changed over time: words in the local language, words in one of several languages, pseudo-words that were “pronounceable”, ten letters with a certain vowel density — and ultimately, any five-letter sequence

Domain-Specific Compression

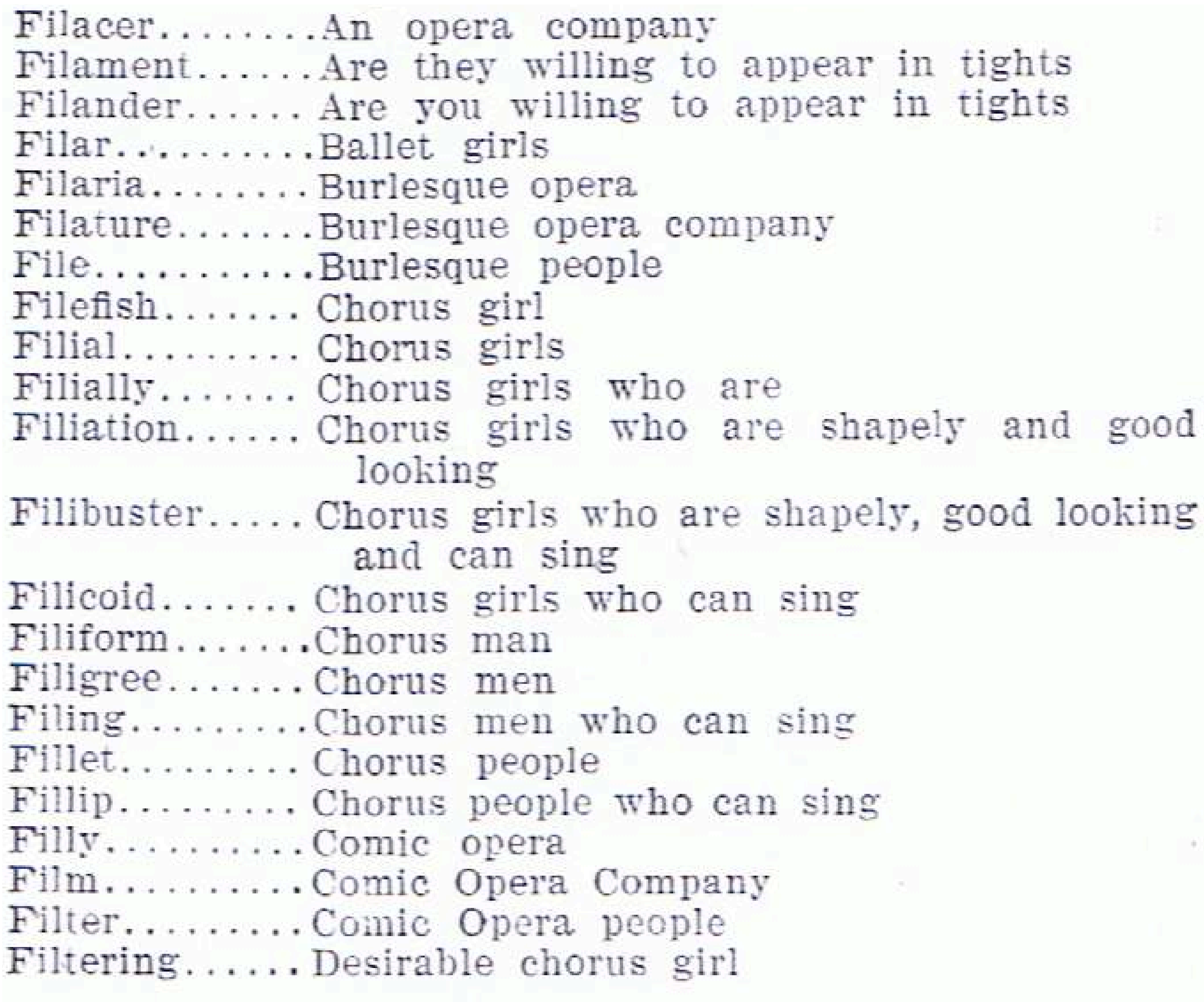
- Many professions had their own code books

- Even explosives manufacturers had their own code

- Example: in the *The Theatrical Cipher Code* (1905), `DISORB` meant `do not want drunkards` and `FILIATION` meant `chorus girls who are shapely and good looking`

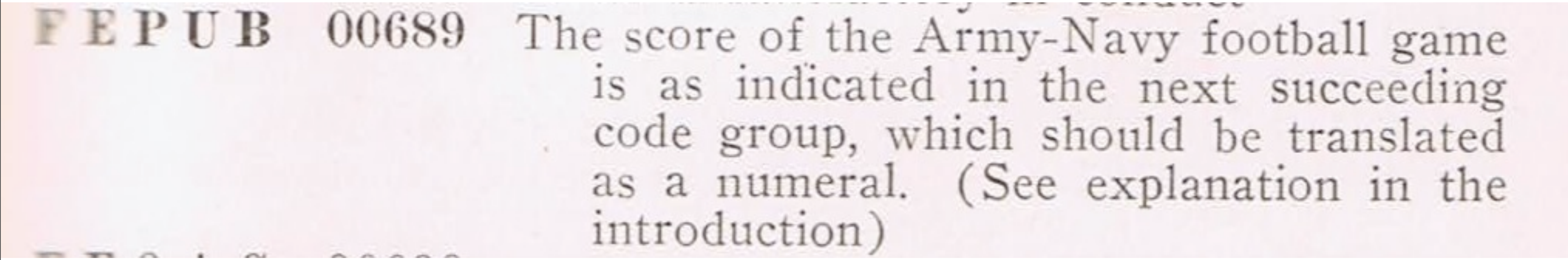
- We still use domain-specific compression: Lempel-Ziv does not work nearly as well as JPEG and MP3 on pictures or audio files

The Theatrical Cipher Code **(1905)**



Filacer	An opera company
Filament	Are they willing to appear in tights
Filander	Are you willing to appear in tights
Filar	Ballet girls
Filaria	Burlesque opera
Filature	Burlesque opera company
File	Burlesque people
Filefish	Chorus girl
Filial	Chorus girls
Filially	Chorus girls who are
Filiation	Chorus girls who are shapely and good looking
Filibuster	Chorus girls who are shapely, good looking and can sing
Filicoid	Chorus girls who can sing
Filiform	Chorus man
Filigree	Chorus men
Filing	Chorus men who can sing
Fillet	Chorus people
Fillip	Chorus people who can sing
Filly	Comic opera
Film	Comic Opera Company
Filter	Comic Opera people
Filtering	Desirable chorus girl

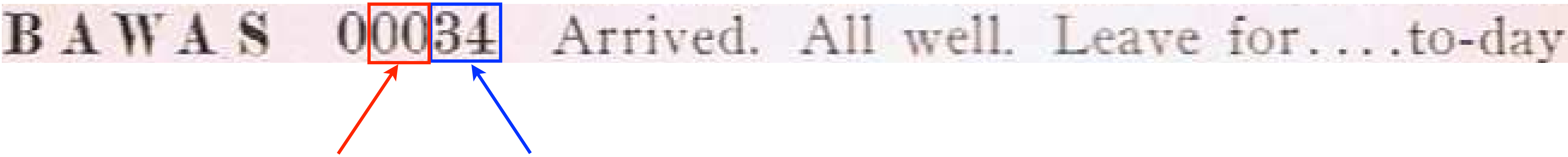
Unofficial Navy Code (1909)



FEPUB 00689 The score of the Army-Navy football game is as indicated in the next succeeding code group, which should be translated as a numeral. (See explanation in the introduction)

For the 2008 game, you would send

FEPUB BAWAS: Army 0, Navy 34



BAWAS 00034 Arrived. All well. Leave for....to-day

Army score

Navy score

Correction

Error Correction

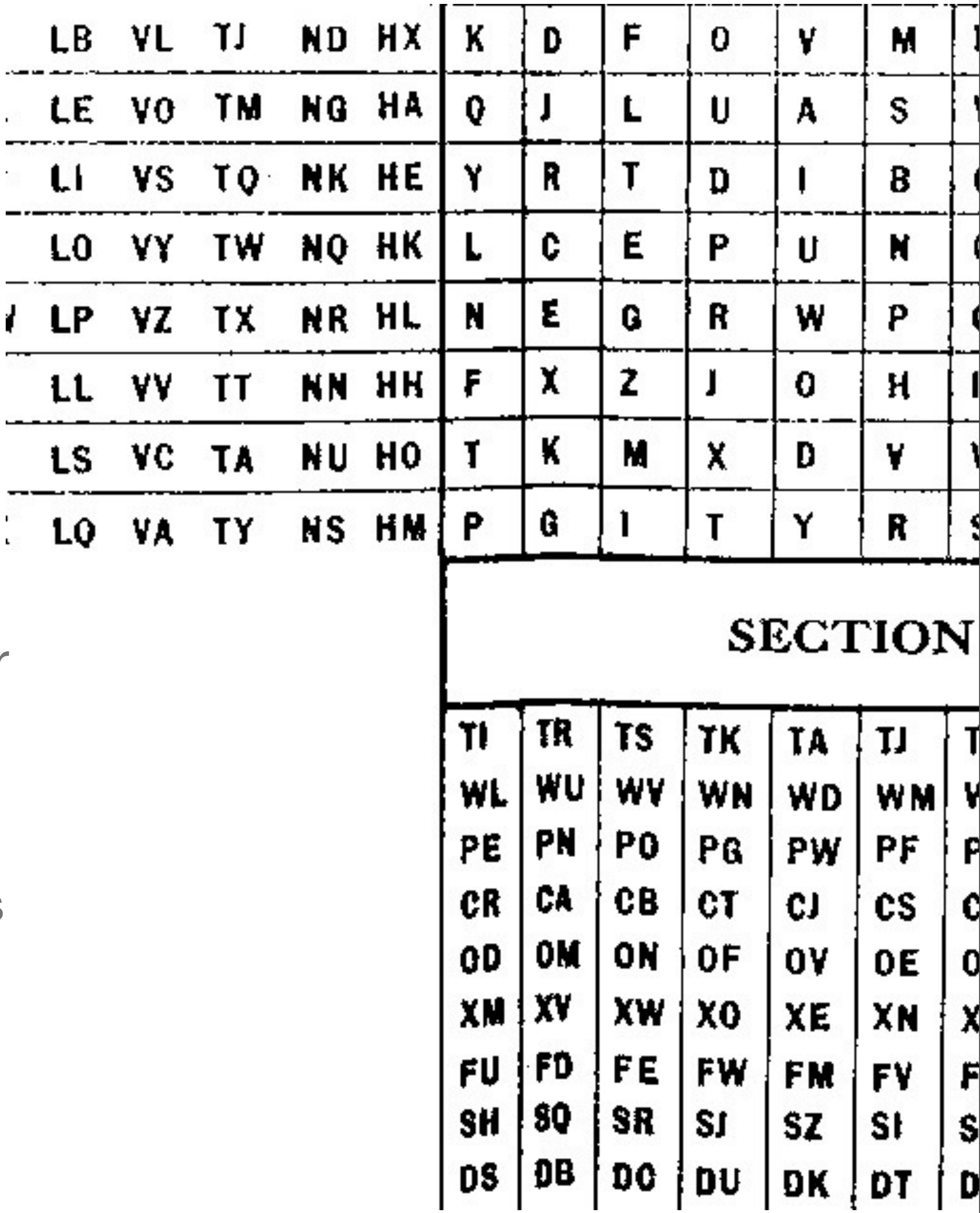
- What about errors during transmission? Errors are link layer-specific, and a given message could be sent over multiple link types
- In the police code, `SUB` is `Vienna`, but `SYB` is `Jerusalem` — and `U` and `Y` are adjacent on the keyboard
- Morse code had its own errors. Consider how .._. (F) could be received:
  - ✦IN (.. _.)
  - ✦ER (. ._.)
  - ✦UE (.._ .)

Techniques

- Terminal indices
- Check digits
- Two-letter differences
- Avoidance of common words
- Mutilation Tables

Mutilation Tables

Look up the first two letters in the upper left, move across to the middle letter, move down to the lower table. Context often permits disambiguation of the possible original words from the various legal possibilities.

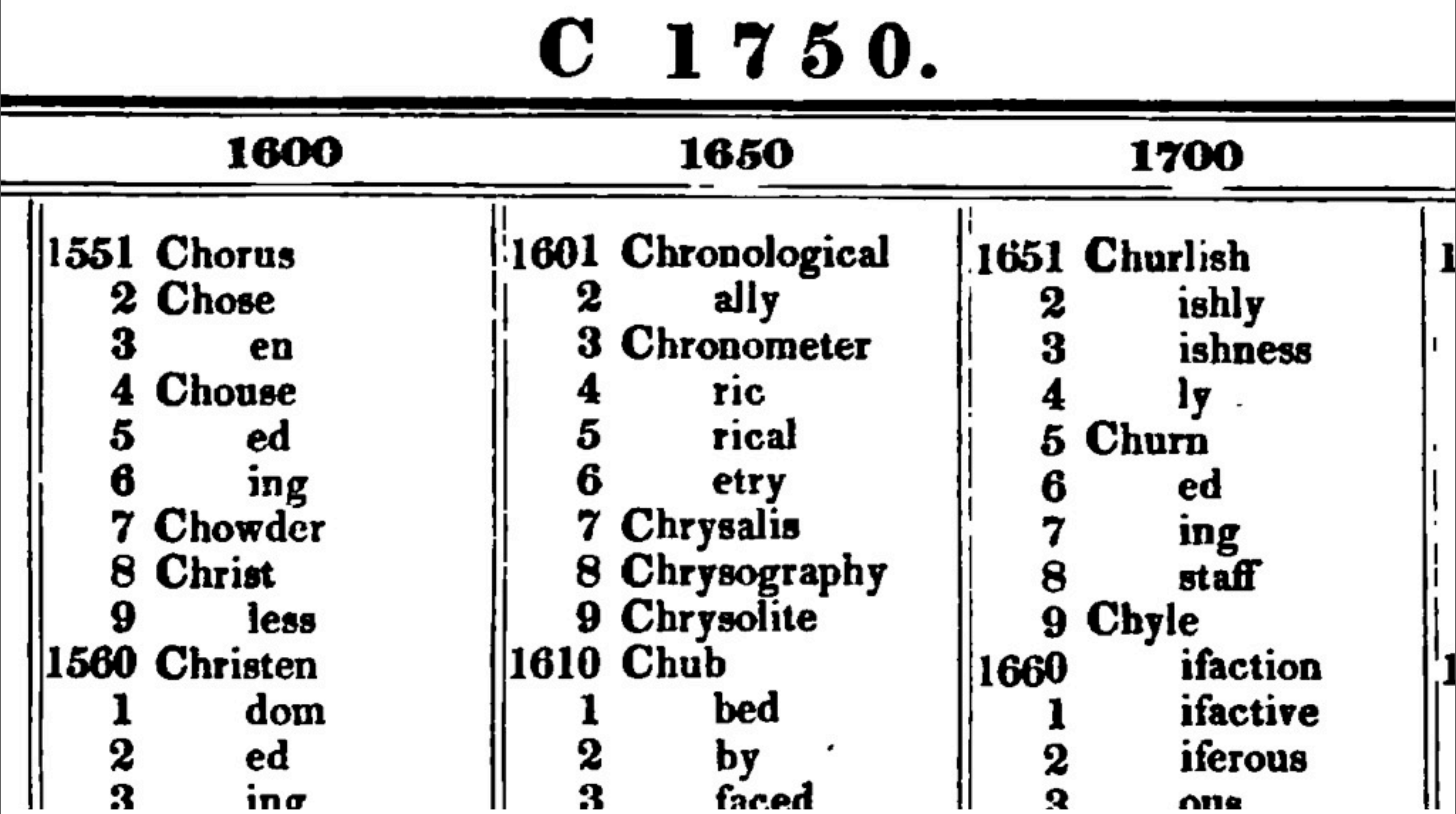


LB	VL	TJ	ND	HX	K	D	F	O	V	M
LE	VO	TM	NG	HA	Q	J	L	U	A	S
LI	VS	TQ	NK	HE	Y	R	T	D	I	B
LO	VY	TW	NQ	HK	L	C	E	P	U	N
LP	VZ	TX	NR	HL	N	E	G	R	W	P
LL	VV	TT	NN	HH	F	X	Z	J	O	H
LS	VC	TA	NU	HO	T	K	M	X	D	V
LQ	VA	TY	NS	HM	P	G	I	T	Y	R

SECTION

TI	TR	TS	TK	TA	TJ
WL	WU	WV	WN	WD	WM
PE	PN	PO	PG	PW	PF
CR	CA	CB	CT	CJ	CS
OD	OM	ON	OF	OV	OE
XM	XV	XW	XO	XE	XN
FU	FD	FE	FW	FM	FV
SH	SQ	SR	SJ	SZ	SI
DS	DB	DO	DU	DK	DT

Confidentiality



C 1750.

1600		1650		1700	
1551	Chorus	1601	Chronological	1651	Churlish
2	Chose	2	ally	2	ishly
3	en	3	Chronometer	3	ishness
4	Chouse	4	ric	4	ly
5	ed	5	rical	5	Churn
6	ing	6	etry	6	ed
7	Chowder	7	Chrysalis	7	ing
8	Christ	8	Chrysography	8	staff
9	less	9	Chrysolite	9	Chyle
1560	Christen	1610	Chub	1660	ifaction
1	dom	1	bed	1	ifactive
2	ed	2	by	2	iferous
3	ing	3	faced	3	ous

The Secret Corresponding Vocabulary (1845)

Add or subtract a prearranged key; monoalphabetic substitution of letter.

Threat Models

On the 1st February, 1870, the telegraph system throughout the United Kingdom passes into the hands of the Government, who will work the lines by Post Office officials. In other words, those who have hitherto so judiciously and satisfactorily managed the delivery of our sealed letters will in future be entrusted also with the transmission and delivery of our open letters in the shape of telegraphic communications, which will thus be exposed not only to the gaze of public officials, but from the necessity of the case must be read by them.

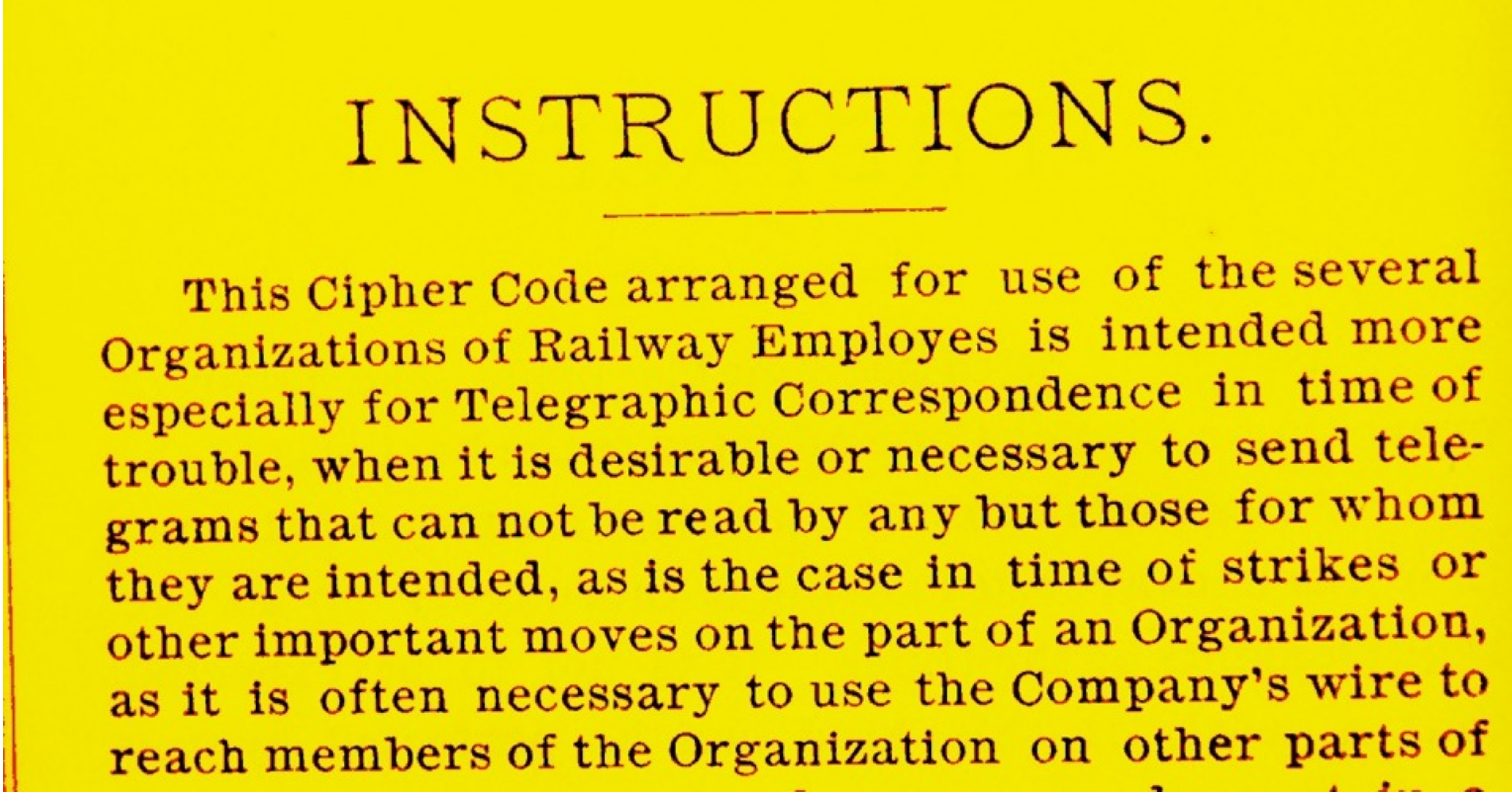
Slater's Telegraphic Code (1870)

Slater’s Telegraph Code (1870-1939)

- Long-lived
- Encode to 5-digit numbers
- Use additives, transpositions of digits, or combinations
- Map result to other code words
- Note: the resulting message was quite expensive: there was no error detection or compression, and the code words were expensive under later tariffs. But the code lasted for almost 70 years.

Bloomer's Commercial Cryptograph: A Telegraph Code and Double Index–Holocryptic Cipher **(1874)**

- Holocryptic: “wholly hidden or secret; spec. of a cipher incapable of being read except by those who have the key” (OED)
- Standard code words, code numbers, and phrases
- Suggestions for additives, transposition of code words, and user-generated two-part code variant
- Different additives could be used for different words (the “holocryptic” part)
- Room for user-created two-part codes (“double index”)

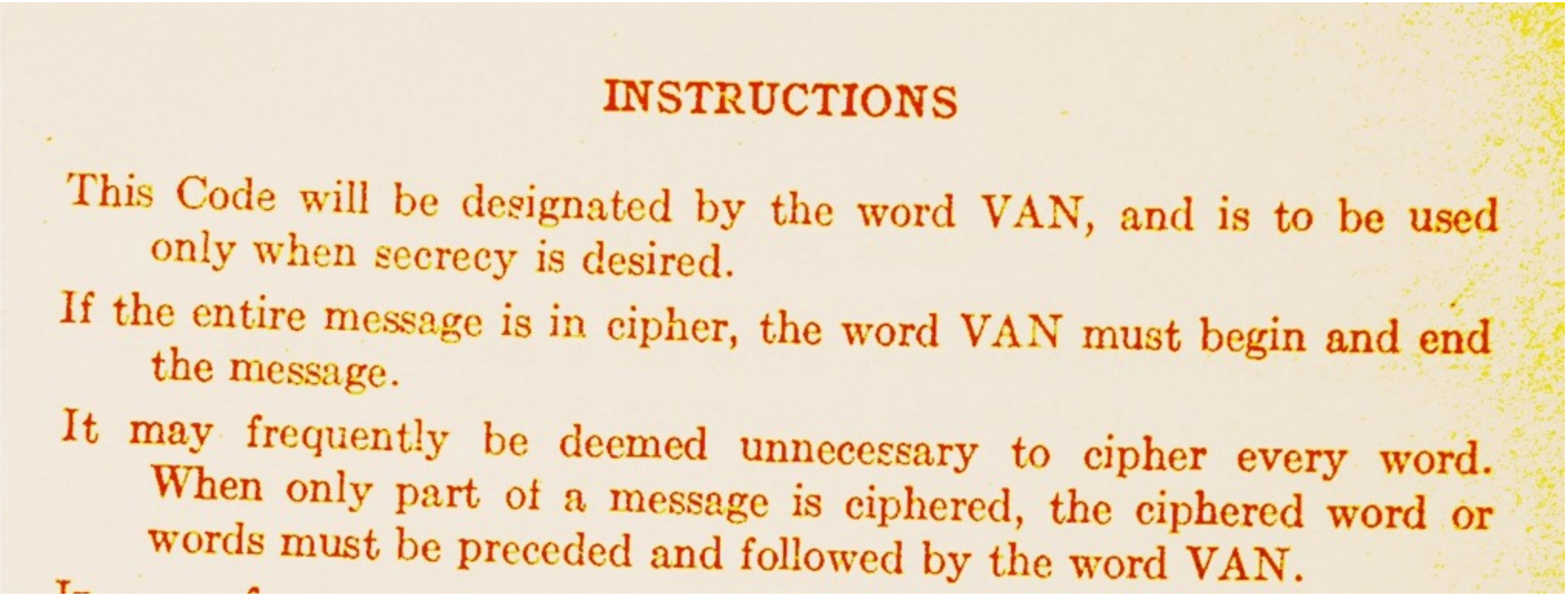


INSTRUCTIONS.

This Cipher Code arranged for use of the several Organizations of Railway Employes is intended more especially for Telegraphic Correspondence in time of trouble, when it is desirable or necessary to send telegrams that can not be read by any but those for whom they are intended, as is the case in time of strikes or other important moves on the part of an Organization, as it is often necessary to use the Company's wire to reach members of the Organization on other parts of

Labor versus Management

Labor had more secure codes...



INSTRUCTIONS

This Code will be designated by the word VAN, and is to be used only when secrecy is desired.

If the entire message is in cipher, the word VAN must begin and end the message.

It may frequently be deemed unnecessary to cipher every word. When only part of a message is ciphered, the ciphered word or words must be preceded and followed by the word VAN.

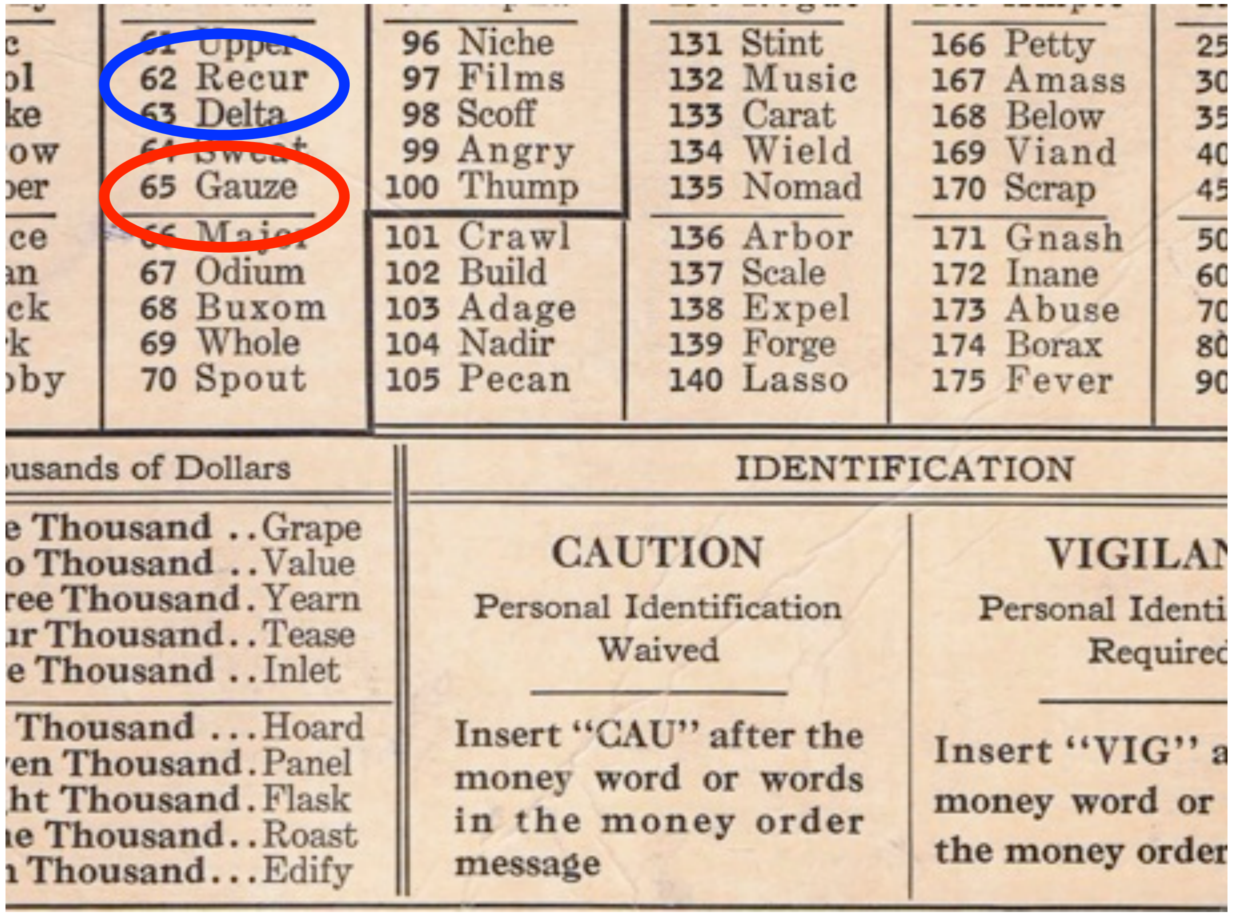
The NY Central's VAN Code (1923)

Governments Were No Better

“When a single key number is used, the number may be alternately added and subtracted. Other methods will readily occur. The use of 50 or 100, while easy to remember, should be avoided.”

U.S. War Department, 1904

Wiring Money: A Two-Part Code (1952)



c	61 Upper	96 Niche	131 Stint	166 Petty	25
ol	62 Recur	97 Films	132 Music	167 Amass	30
ke	63 Delta	98 Scoff	133 Carat	168 Below	35
ow	64 Sweat	99 Angry	134 Wield	169 Viand	40
oer	65 Gauze	100 Thump	135 Nomad	170 Scrap	45
ce	66 Major	101 Crawl	136 Arbor	171 Gnash	50
an	67 Odium	102 Build	137 Scale	172 Inane	60
ck	68 Buxom	103 Adage	138 Expel	173 Abuse	70
k	69 Whole	104 Nadir	139 Forge	174 Borax	80
oby	70 Spout	105 Pecan	140 Lasso	175 Fever	90

usands of Dollars

e Thousand	..Grape
o Thousand	..Value
ree Thousand	.Yearn
ur Thousand	..Tease
e Thousand	..Inlet
Thousand	...Hoard
en Thousand	.Panel
ht Thousand	.Flask
ne Thousand	..Roast
n Thousand	...Edify

IDENTIFICATION

CAUTION

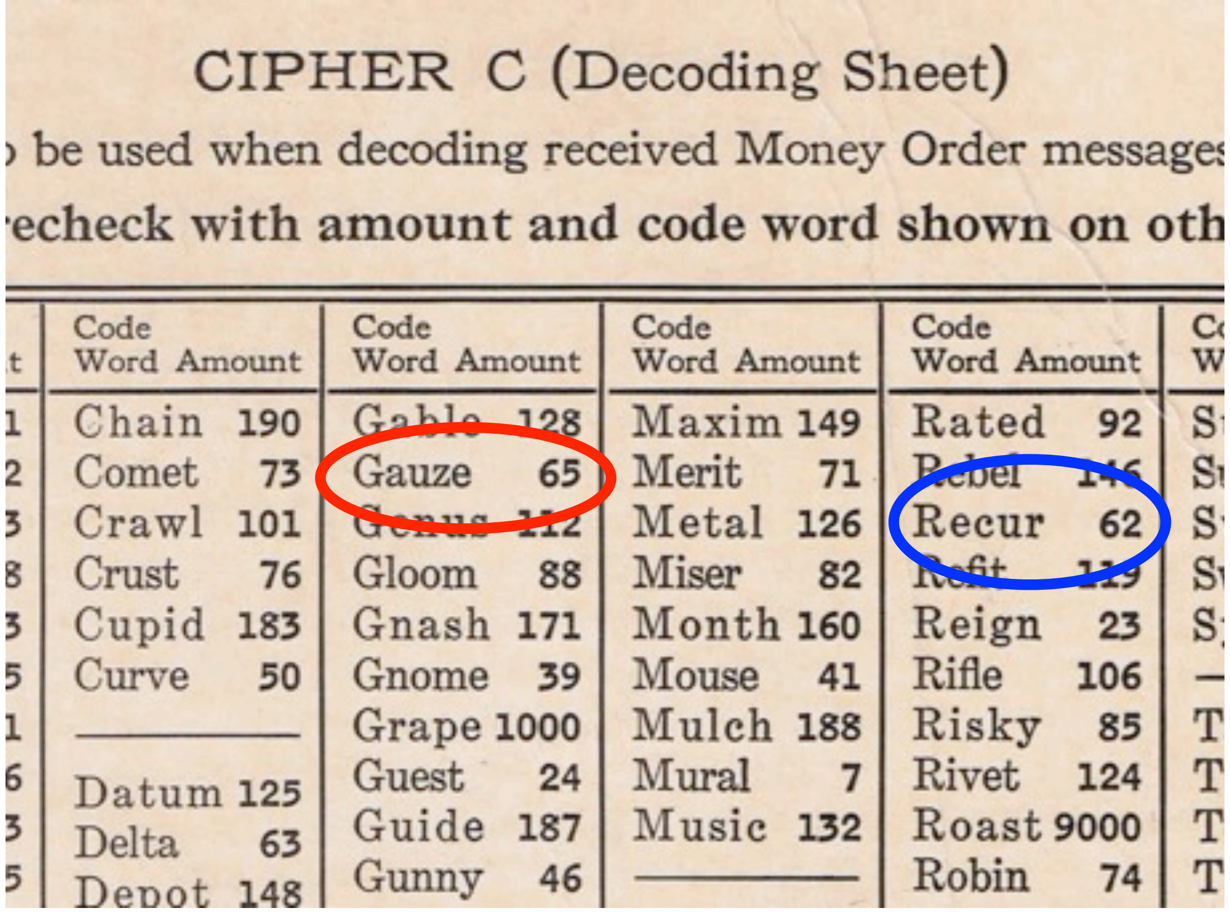
Personal Identification
Waived

Insert "CAU" after the money word or words in the money order message

VIGILAN

Personal Identi
Required

Insert "VIG" a
money word or
the money order



CIPHER C (Decoding Sheet)

be used when decoding received Money Order message

echeck with amount and code word shown on oth

t	Code Word	Amount	Code Word	Amount	Code Word	Amount	Code Word	Amount	Co W
1	Chain	190	Gable	128	Maxim	149	Rated	92	S
2	Comet	73	Gauze	65	Merit	71	Rebel	146	S
3	Crawl	101	Genus	112	Metal	126	Recur	62	S
8	Crust	76	Gloom	88	Miser	82	Refit	119	S
3	Cupid	183	Gnash	171	Month	160	Reign	23	S
5	Curve	50	Gnome	39	Mouse	41	Rifle	106	—
1	—		Grape	1000	Mulch	188	Risky	85	T
6	Datum	125	Guest	24	Mural	7	Rivet	124	T
3	Delta	63	Guide	187	Music	132	Roast	9000	T
5	Depot	148	Gunny	46	—		Robin	74	T

Comprehension

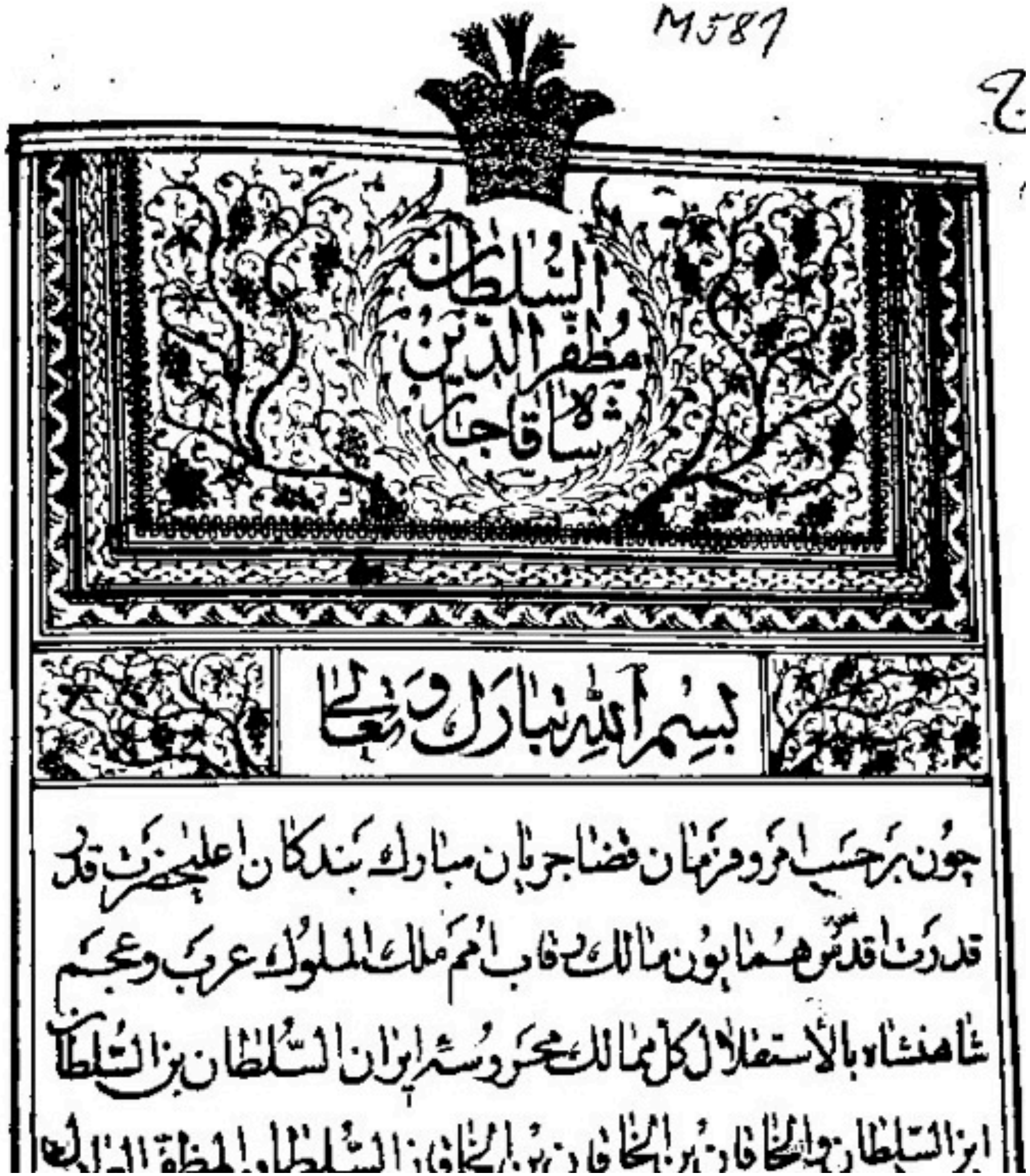
Comprehension

- “A code reflects the world at a particular instant, and as the world moves on it outmodes the code. New products, new ways of doing things, new political or economic facts begin to make its vocabulary old-fashioned.” (Kahn)
- Code books present a picture of a given era
- Code books could also be used for translation

A Bygone Age

- “Marriage has been arranged between _____” (*Unicode*, 1897)
- “Will lunch with you today” (*Unicode)*
- “Roman Catholic intrigue” (*China Inland Mission Private Telegraph Code*, 1907)
- “Send women on shore to wash” (*Popham’s Naval Signal Code*, under “Military and Technical Terms”, 1816)
- Professions: “Castle-keeper” (*International Police Telegraph Code*,1930)

Cultural Norms: An Illuminated Persian Government Codebook (1901)



M587

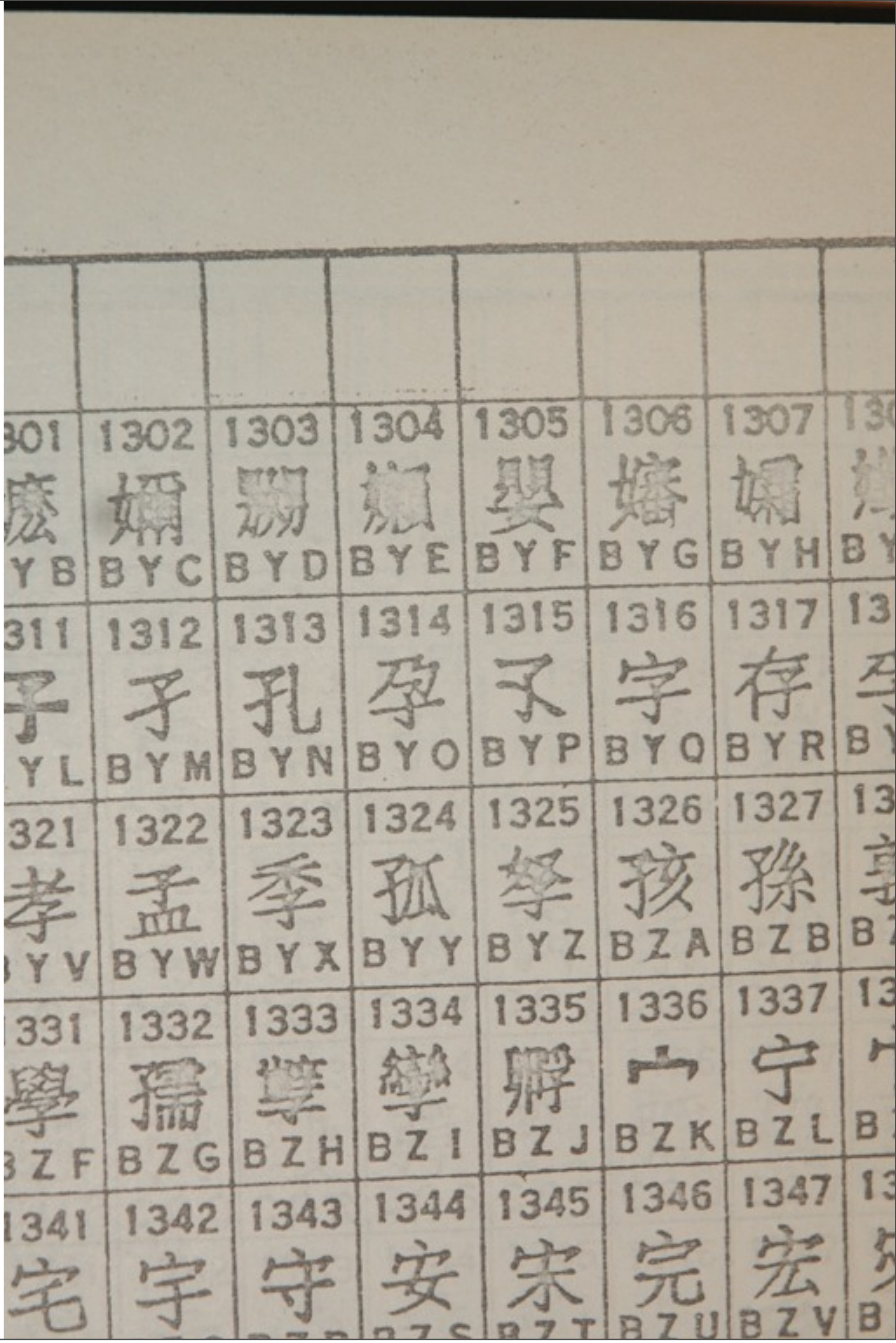
السلطان مظفر الدین شاه قاجار

بسم الله تبارک و تعالی

چون برحسب امر و فرمان قضا جریان مبارک بندگان اعلیحضرت قدر
قدرت اقدس همایون مالک رقاب امم ملک الملوک عرب و عجم
شاهنشاه بالاستقلال کل ممالک محروسه ایران السلطان بن السلطان
ابن السلطان و الخاقان بن الخاقان بن الخاقان السلطان المظفر الدین

Sending Chinese Characters

- 4-digit/3-letter link-layer encoding for each Chinese character
- Widely used in China until about 10 years ago — faxes and cell phones have taken over
- Code points are still used today for names on official forms: dialect-independent, unambiguous, etc.



301 嬤 YB	1302 嬬 BYC	1303 嬲 BYD	1304 嬾 BYE	1305 嬰 BYF	1306 嬸 BYG	1307 孎 BYH	13 B
311 子 YL	1312 孑 BYM	1313 孔 BYN	1314 孕 BYO	1315 孓 BYP	1316 字 BYQ	1317 存 BYR	13 B
321 孝 YV	1322 孟 BYW	1323 季 BYX	1324 孤 BYY	1325 孥 BYZ	1326 孩 BZA	1327 孫 BZB	13 B
331 學 ZF	1332 孺 BZG	1333 孼 BZH	1334 孿 BZI	1335 孵 BZJ	1336 宀 BZK	1337 宁 BZL	13 B
1341 宅	1342 宇	1343 守	1344 安 BZS	1345 宋 BZT	1346 完 BZU	1347 宏 BZV	13 B

Copyright Infringement

- The U.S. hadn’t signed the Berne Convention; books weren’t protected here unless printed here first. Some code books were widely pirated.
    - ✦British publishers sometimes printed the first edition in the U.S. to avoid that
    - ✦Note — pirate editions couldn’t be imported into the British empire
- The code words themselves were valuable; those were pirated, too

Summary

Most Functions Existed at All Layers

	Link	Codebook	Plaintext
Compression	Morse code elements	Careful phrase selection	Restricted word choice; sentence fragments
Correction	Different links have different error properties; read-back	Mutilation tables; terminal indices; Hamming distance; check digits	Use code words for numbers
Confidentiality	Avoid exposed links (i.e., radio; other countries' wires)	Superencryption; secret codebooks	Semantically combine fields

There were generally tradeoffs in convenience, performance, efficiency, etc.

Parting Thoughts

- Telegraph codebooks were used in Australia until at least 1972, and in China until around 2000

- What we do today is the evolution of what was done then

- Huffman didn't invent compression; Hamming didn't invent error correction; NIST didn't invent encryption

- (Draft paper at papers/codebooks.pdf on my web page.)

Acknowledgments

- Jim Reeds
- Hang Zhao, Seung Geol Choi, Evelyn Guzman, Malek Ben Salem, Arezu Moghadam, and Ted Lemon
- The research libraries and librarians of the world
- Google Books

Compression, Correction, Confidentiality, and Comprehension

Steven M. Bellovin
smb@cs.columbia.edu
http://www.cs.columbia.edu/~smb

+1 212-939-7149
Department of Computer Science
Columbia University

cc BY NC