Compression, Correction, Confidentiality, and Comprehension: A Look at Telegraph Codes

Steven M. Bellovin
Columbia University
smb@cs.columbia.edu

Abstract

Telegraph codes are a more-or-less forgotten part of technological history. In their day, though, they were ubiquitous and sophisticated. They also laid the groundwork for many of today's communications technologies, including encryption, compression, and error correction. Beyond that, reading them provides a snapshot into culture. We look back, describing them in modern terms, and noting some of the tradeoffs considered.

1 Introduction

Most cryptologists have heard of telegraph codes. Often, though, our knowledge is cursory. We've forgotten what we read in Kahn [9], and perhaps remember little more than the basic concept: a word, phrase, or sentence is represented by a single codeword. In fact, telegraph codes were far more sophisticated, and laid the groundwork for many later, fundamental advances.

Looked at analytically, telegraph codes fulfilled four primary functions: compression, correction, confidentiality, and comprehension. Beyond that, they offer a window into the past: the phrases they can be used to represent give insight into daily lives of the time.

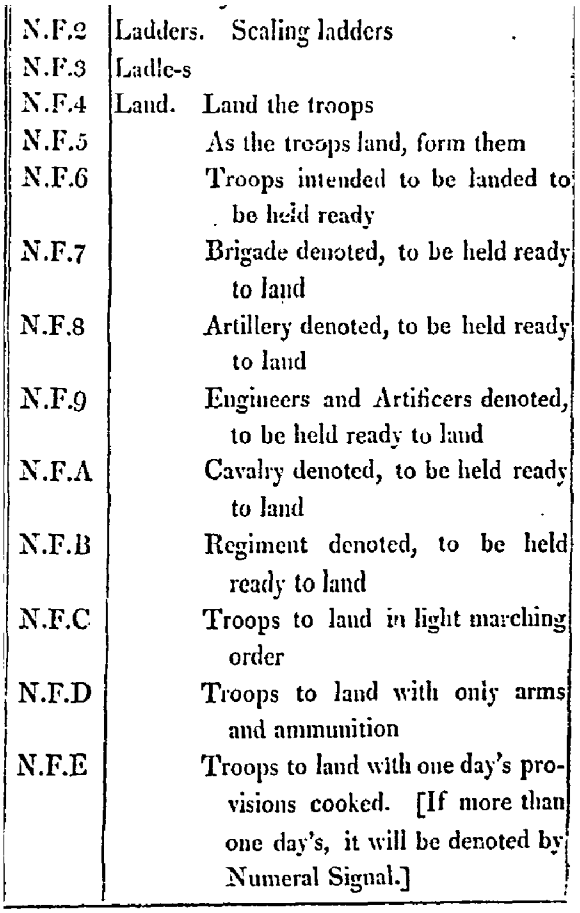
This paper, based primarily on my own small collection of codebooks, illustrates some of these points. For typographical simplicity, I have written codewords **LIKE THIS**, while plaintext is written *this way.*

2 Compression

Compression was the original goal of telegraph codes. Early trans-Atlantic telegrams were *extremely* expensive — $100 for twenty words in 1866 [8] — so brevity was very important.

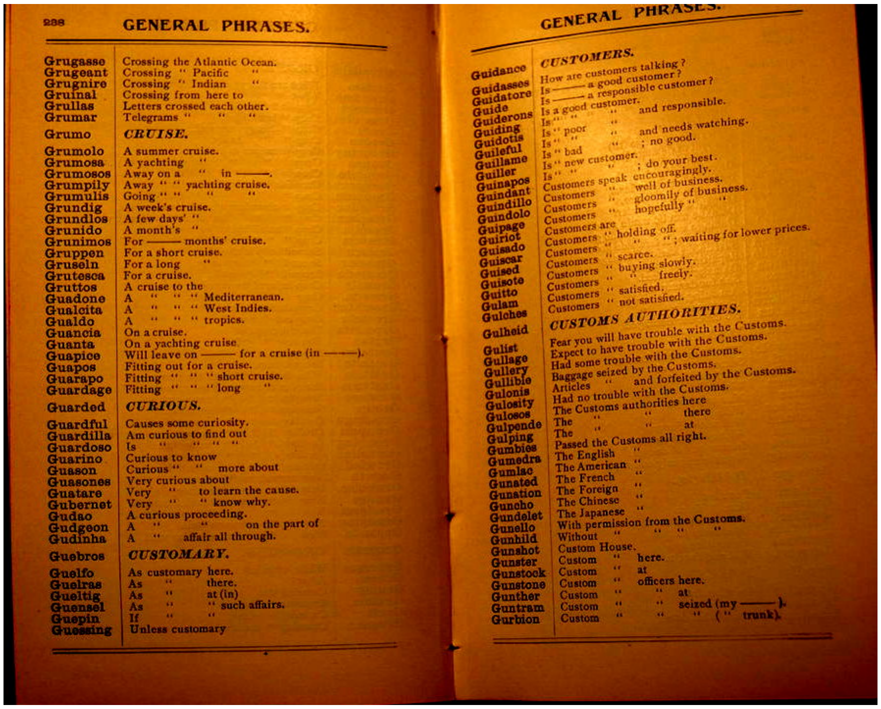
Early telegraph codes had two ancestors, codes for semaphore networks and naval signaling [9]. The constraint in the latter case was not so much cost (though rifling through a collection of flags would not have been quick); rather, the issue was limited space on a ship's rigging for the flags. Early naval codes conveyed meaning by a combination of flag and location. The vocabulary was very limited; it was not possible to send arbitrary messages in such schemes. Later, *numerary codes* were introduced, where a set of flags representing digits were used to indicate an entry in a signal book. The first such system is often attributed to Admiral Bertrand-François Mahé de la Bourdonnais [14, 6]; however, it was not adopted, possibly because he was of insufficiently noble birth. A number of British admirals adopted and adapted this scheme. Sir Charles Knowles devised a matrix system for indicating digits; a pair of flags, one over the other, would select a matrix cell for a given value. Later, Admiral Richard Lord Howe, probably with the assistance of Captain Richard Kempenfelt (who was familiar with Mahé de la Bourdonnais's work), devised a longer and better signal book. His first version also used a tabular scheme. Amusingly enough from a computer scientist's perspective, he used a 16×16 matrix: one byte!

The most influential early numerary code was devised by Sir Home Popham in 1803 [6, 14, 15, 23]; it included concise signals for such phrases as "Troops to land with one day's provisions cooked" (Figure 1).



N.F.2	Ladders.	Scaling ladders
N.F.3	Ladles	
N.F.4	Land.	Land the troops
N.F.5		As the troops land, form them
N.F.6		Troops intended to be landed to be held ready
N.F.7		Brigade denoted, to be held ready to land
N.F.8		Artillery denoted, to be held ready to land
N.F.9		Engineers and Artificers denoted, to be held ready to land
N.F.A		Cavalry denoted, to be held ready to land
N.F.B		Regiment denoted, to be held ready to land
N.F.C		Troops to land in light marching order
N.F.D		Troops to land with only arms and ammunition
N.F.E		Troops to land with one day's provisions cooked. [If more than one day's, it will be denoted by Numeral Signal.]

Figure 1: Popham's naval code.



238 GENERAL PHRASES.

Grugasse	Crossing the Atlantic Ocean.
Grugeant	Crossing " Pacific "
Grugnire	Crossing " Indian "
Gruinal	Crossing from here to
Grullas	Letters crossed each other.
Grumar	Telegrams " " "
Grumo	*CRUISE.*
Grumolo	A summer cruise.
Grumosa	A yachting "
Grumosos	Away on a " in ———.
Grumpily	Away " " yachting cruise.
Grumulis	Going " " " "
Grundig	A week's cruise.
Grundlos	A few days' "
Grunido	A month's "
Grunimos	For ——— months' cruise.
Gruppen	For a short cruise.
Gruseln	For a long "
Grutesca	For a cruise.
Gruttos	A cruise to the
Guadone	A " " " Mediterranean.
Gualcita	A " " " West Indies.
Gualdo	A " " " tropics.
Guancia	On a cruise.
Guanta	On a yachting cruise
Guapice	Will leave on ——— for a cruise (in ———).
Guapos	Fitting out for a cruise.
Guarapo	Fitting " " " short cruise.
Guardage	Fitting " " " long "
Guarded	*CURIOUS.*
Guardful	Causes some curiosity.
Guardilla	Am curious to find out
Guardoso	Is " " " "
Guarino	Curious to know
Guason	Curious " " more about
Guasones	Very curious about
Guatare	Very " to learn the cause.
Gubernet	Very " " know why.
Gudao	A curious proceeding.
Gudgeon	A " " on the part of
Gudinha	A " affair all through.
Guebros	*CUSTOMARY.*
Guelfo	As customary here.
Guelras	As " there.
Gueltig	As " at (in)
Guensel	As " " such affairs.
Guepin	If " "
Guessing	Unless customary

GENERAL PHRASES.

Guidance	*CUSTOMERS.*
Guidasses	How are customers talking?
Guidatore	Is ——— a good customer?
Guide	Is ——— a responsible customer?
Guiderons	Is a good customer.
Guiding	Is " " " and responsible.
Guidotis	Is " poor " and needs watching.
Guileful	Is " " " ; no good.
Guillame	Is " bad "
Guiller	Is " new customer.
Guinapos	Is " " " ; do your best.
Guindant	Customers speak encouragingly.
Guindillo	Customers " well of business.
Guindolo	Customers " gloomily of business.
Guipage	Customers " hopefully " "
Guiriot	Customers are
Guisado	Customers " holding off.
Guiscar	Customers " " " ; waiting for lower prices.
Guised	Customers " scarce.
Guisote	Customers " buying slowly.
Guitto	Customers " " freely.
Gulam	Customers " satisfied.
Gulches	Customers " not satisfied.
Gulheid	*CUSTOMS AUTHORITIES.*
Gulist	Fear you will have trouble with the Customs.
Gullage	Expect to have trouble with the Customs.
Gullery	Had some trouble with the Customs.
Gullible	Baggage seized by the Customs.
Gulonis	Articles " and forfeited by the Customs.
Gulosity	Had no trouble with the Customs.
Gulosos	The Customs authorities here
Gulpende	The " " there
Gulping	The " " at
Gumbies	Passed the Customs all right.
Gumedra	The English "
Gumlac	The American "
Gunated	The French "
Gunation	The Foreign "
Guncho	The Chinese "
Gundelet	The Japanese "
Gunello	With permission from the Customs.
Gunhild	Without " " " "
Gunshot	Custom House.
Gunster	Custom " here.
Gunstock	Custom " at
Gunstone	Custom " officers here.
Gunther	Custom " " at
Guntram	Custom " " seized (my ———).
Gurbion	Custom " " " (" trunk).

Figure 2: Some sample pages from the 1896 *Atlas Universal Travelers' and Business Telegraphic Cipher Code.*

More importantly, it provide signals for various individual parts of speech. [6] likens it to "the step from a "Traveller's Manual of Conversation" to a dictionary of the language". Popham's code included the concept of parameters; thus, the previous phrase could be modified: "If more than one day's, it will be denoted by Numeral Signal." Nelson's famous signal "England expects that every man will do his duty" was sent using this's system [14]. It was eventually adopted by the Admiralty as the standard signal book [44].

Frederick Marryat produced a *Code of Signals for the Merchant Service* in 1817. It assigned a 4-digit code to each sentence and to each individual merchant ship [19]. By 1828, there was even a codebook for yachts and pleasure boats [51].

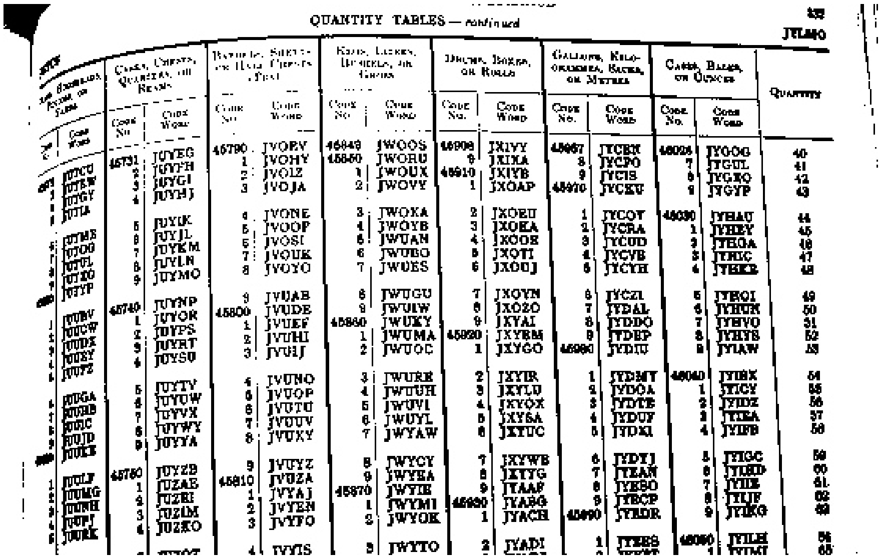
Further details on the evolution of naval signals are beyond the scope of this work; those interested should see [6, 14, 15, 23].

Telegraph codes drew on this rich history. Many codebooks were aimed at shipping and traveling. Figure 2 shows a typical example of this genre. By the time this particular code was issued, a great deal of effort had gone into specialized phrases. Thus, **GULLIBLE** meant *Baggage seized by the Customs*, while **GURBION** meant *Custom House officers seized my trunk.*

Some codebooks incorporated domain-specific information. Thus, when Charles A. Stoneham & Co., a mining stock brokerage firm, issued its own codebook in 1910, it had words such as **REVERE** meaning *Wires being down, your telegram did not reach us in time to transact any business today, and as your orders are good for the week, we will try to execute tomorrow.* There were also specific code words for gold fields, mining companies, etc.

Domain-specific compression is at least as important today. MP3 and JPEG compression of sound and picture files is far more effective than, say, simple Lempel-Ziv compression would be. Informal comparisons show that high-quality JPEG images taken from "raw" camera files are 30–40% smaller; simple Lempel-Ziv compression achieved no more than 5% improvement in file size.

Referring back to Figure 2, note that some messages are parameterized. Thus, one might send **GUIDE**



QUANTITY TABLES — *continued*

JULMO

[illegible]		Cases, Chests, Quarters, or Reams		Barrels, Sheets, or Half Chests, Feet		Kegs, Lockers, Bushels, or Gross		Drums, Boxes, or Rolls		Gallons, Kilogrammes, Sacks, or Metres		Cases, Bales, or Ounces		Quantity
Code No.	Code Word	Code No.	Code Word	Code No.	Code Word	Code No.	Code Word	Code No.	Code Word	Code No.	Code Word	Code No.	Code Word	
[illegible]	JUTCU	45731	JUYEG	45790	JVORY	45849	JWOOS	45908	JXIVY	45967	JYCEN	46026	JYGOG	40
[illegible]	JUTEW	2	JUYFH	1	JVOHY	45850	JWORU	9	JXIXA	8	JYCFO	7	JYGUL	41
[illegible]	JUTGY	3	JUYGI	2	JVOLZ	1	JWOUX	45910	JXIYB	9	JYCIS	8	JYGXO	42
[illegible]	JUTIA	4	JUYHJ	3	JVOJA	2	JWOVY	1	JXOAP	45970	JYCKU	9	JYGYP	43
[illegible]	JUTME	5	JUYIK	4	JVONE	3	JWOKA	2	JXOEU	1	JYCOY	46030	JYHAU	44
[illegible]	JUTOO	6	JUYJL	5	JVOOP	4	JWOYB	3	JXOEA	2	JYCRA	1	JYHEY	45
[illegible]	JUTUL	7	JUYKM	6	JVOSI	5	JWUAN	4	JXOOS	3	JYCUD	2	JYHGA	46
[illegible]	JUTZO	8	JUYLN	7	JVOUK	6	JWUBO	5	JXOTI	4	JYCVB	3	JYHIC	47
[illegible]	JUTYP	9	JUYMO	8	JVOYO	7	JWUES	6	JXOUJ	5	JYCYH	4	JYHER	48
[illegible]	JUUBV	45740	JUYNP	9	JVUAB	8	JWUGU	7	JXOYN	6	JYCZI	5	JYHOI	49
[illegible]	JUUCW	1	JUYOR	45800	JVUDE	9	JWUIW	8	JXOZO	7	JYDAL	6	JYHUK	50
[illegible]	JUUDX	2	JUYPS	1	JVUEF	45860	JWUKY	9	JXYAI	8	JYDDO	7	JYHVO	51
[illegible]	JUUEY	3	JUYRT	2	JVUHI	1	JWUMA	45920	JXYBM	9	JYDEP	8	JYHYS	52
[illegible]	JUUFZ	4	JUYSU	3	JVUIJ	2	JWUOC	1	JXYGO	45980	JYDIU	9	JYIAW	53
[illegible]	JUUGA	5	JUYTV	4	JVUNO	3	JWURE	2	JXYIR	1	JYDMY	46040	JYIBX	54
[illegible]	JUUHB	6	JUYUW	5	JVUOP	4	JWUUH	3	JXYLU	2	JYDOA	1	JYICY	55
[illegible]	JUUIC	7	JUYVX	6	JVUTU	5	JWUVI	4	JXYOX	3	JYDTE	2	JYIDZ	56
[illegible]	JUUJD	8	JUYWY	7	JVUUV	6	JWUYL	5	JXYSA	4	JYDUF	3	JYIEA	57
[illegible]	JUUKE	9	JUYYA	8	JVUXY	7	JWYAW	6	JXYUC	5	JYDXI	4	JYIFB	58
[illegible]	JUULF	45750	JUYZB	9	JVUYZ	8	JWYCY	7	JXYWB	6	JYDYJ	5	JYIGC	59
[illegible]	JUUMG	1	JUZAE	45810	JVUZA	9	JWYEA	8	JXYYG	7	JYEAN	6	JYIHD	60
[illegible]	JUUNH	2	JUZBI	1	JVYAJ	45870	JWYIE	9	JYAAF	8	JYEBO	7	JYIIE	61
[illegible]	JUUPJ	3	JUZIM	2	JVYEN	1	JWYMI	45930	JYABG	9	JYECP	8	JYIJF	62
[illegible]	JUURK	4	JUZKO	3	JVYFO	2	JWYOK	1	JYACH	45990	JYEDR	9	JYIKG	63
		[illegible]	[illegible]	4	JVYIS	3	JWYTO	2	JYADI	1	JYEES	46050	JYILH	64
												1	JYIMI	65

Figure 3: Part of a quantity table from the 1936 *A B C Telegraphic Code, Seventh Edition* [43].

RICANTI to ask *Is **RICANTI** a responsible customer?* (Note the potential for confusion: is Ricanti a proper name? In fact, it is the code word for the *Bank of Ireland*. This issue is discussed further in Section 3.)

Sometimes, compression was implicit. Figure 3 shows that **JYGUL** (or **46026**) could stand for *41 Cases*, *41 Bales*, or *41 Ounces*. Presumably, the recipient would know what was meant. On the other hand, there were typically distinct code words for amounts of money in dollars, pounds sterling, francs, etc.

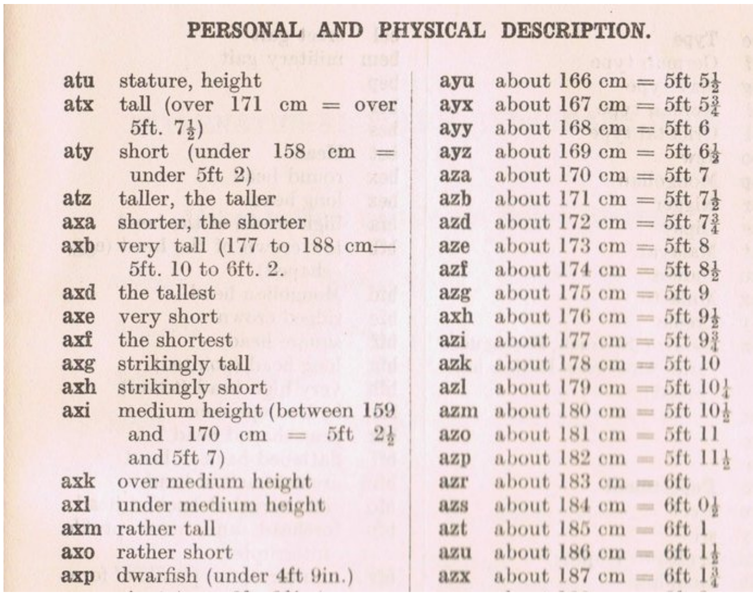
Compression could be taken to extremes. One wonders how ofen users of the 1920 *ABC Telegraphic Code, Sixth Edition* ever sent **ENBET** (*Captain is insane*) or **PAASG** (*Arrived here (at —), encountered a severe gale and heavy seas, which carried away boats and wheel, stanchions and bulwarks, broke mast and jib-boom, all sails gone.*[1] (We note that the first of these phrases persisted into the 7th Edition, from 1936, but the second did not. Was there a greater incidence of crazy captains than bad weather?)

As telegraphy evolved, compression efficiency was no longer measured in characters but in money. What was actually charged by the telegraph companies was the important matter, and code compilers were quick to find loopholes. Instead of sending **MAGFD YHFJU** for *Delivered in time to San Francisco*, could one send **MAGFDYHFJU** and thus be charged for a single word? What *is* a word? In 1903, international regulations defined a "word" as ten characters or less "capable of pronunciation according to the usage of one of the folloiwng languages: German, English, Spanish, French, Dutch, Italian, Portuguese, or Latin". This gave rise to things like the 1907 *Pantelegraphy Simplex Translating and Check Card*, which allowed digits or pairs of digits to be encoded as consonant-vowel pairs, with alternate forms in the name of euphony. When the regulations loosened in 1929, to eliminate the pronouncability requirement but to impose a vowel density stamdard, code makers adapted accordingly.

3 Correction

Especially towards the end of the codebook period, a tremendous amount of effort went into error detection and correction. Errors could be costly, in time, money, or both, and the encoding process removed a lot of redundancy. Consider the poor constable who typed **AXF** instead of **AXG** (Figure 4) — and note that **F**

[1] While I certainly recalled Kahn's reprint [9, p. 851] of the classic July 28, 1934 *New Yorker* essay on amusing code words while I was writing this, I did not, in fact, have it available at the time. I later realized that Kahn misidentified the code as the *Acme*; it's actually the *A B C Sixth Edition* — with some errors!



PERSONAL AND PHYSICAL DESCRIPTION.

atu	stature, height
atx	tall (over 171 cm = over 5ft. 7½)
aty	short (under 158 cm = under 5ft 2)
atz	taller, the taller
axa	shorter, the shorter
axb	very tall (177 to 188 cm— 5ft. 10 to 6ft. 2.
axd	the tallest
axe	very short
axf	the shortest
axg	strikingly tall
axh	strikingly short
axi	medium height (between 159 and 170 cm = 5ft 2½ and 5ft 7)
axk	over medium height
axl	under medium height
axm	rather tall
axo	rather short
axp	dwarfish (under 4ft 9in.)
ayu	about 166 cm = 5ft 5½
ayx	about 167 cm = 5ft 5¾
ayy	about 168 cm = 5ft 6
ayz	about 169 cm = 5ft 6½
aza	about 170 cm = 5ft 7
azb	about 171 cm = 5ft 7½
azd	about 172 cm = 5ft 7¾
aze	about 173 cm = 5ft 8
azf	about 174 cm = 5ft 8½
azg	about 175 cm = 5ft 9
axh	about 176 cm = 5ft 9½
azi	about 177 cm = 5ft 9¾
azk	about 178 cm = 5ft 10
azl	about 179 cm = 5ft 10¼
azm	about 180 cm = 5ft 10½
azo	about 181 cm = 5ft 11
azp	about 182 cm = 5ft 11½
azr	about 183 cm = 6ft
azs	about 184 cm = 6ft 0½
azt	about 185 cm = 6ft 1
azu	about 186 cm = 6ft 1½
azx	about 187 cm = 6ft 1¾

Figure 4: Part of a police code for describing suspects.

and **G** are adjacent on the keyboard, though it isn't clear that it ever would have been typed as opposed to hand-written and sent in Morse code. (This code was rather late for such errors. It authors were apparently more concerned with economy: three codewords could be combined into a single telegraph word for billing purposes.)

A number of different techniques were used. *Mutilation tables* are perhaps the most interesting. A chart (Figure 5) shows the possible middle letters of a codeword, indexed by the initial two and final two letters that can produce that letter. Consider the chart shown, and a received code word **ZNBAB**. This is an impossible value; if the first two letters are **ZN** and the last two **AB**, the middle letter must be **N**. The error could be in any of the three sections, leading to five possible correct values; the instructions suggest looking at the semantics of the decoded value to reconstruct the proper plaintext.

As a complement to mutilation tables, terminal indices were sometimes provided. These were indices alphabetized by the last two letters of the codeword, and were used when it was suspected that the beginning of the word had been corrupted in transmission.

Numerical data data was particularly sensitive. The semantic difference between, say, "1,000" and "9,000' is small, so it is harder to recognize errors from context; nevertheless, the business difference can be great. Accordingly, code makers adopted check digits or letters (Figure 6). These are what today would be known as checksums over the plaintext, and provide at least error detection. One code, the 1929 *New Standard Code*, had a separate set of mutilation tables for numeric data. (In fact, that understates the cleverness of their solution: numeric data was a special case of "subsidiary tables", which were used to encode not just numbers of various types (currency, dimensions, etc.) but also things like repeated dates and markings: **CD** or **AX** might mean *best possible shipment January*, **CM** or **BU** *best possible shipment August*, etc., all concatenated into a single codeword, and followed by a check letter.)

Check letters are notable because they operate on the plaintext, and thus can help with encoding errors. More common techniques dealt with transmission errors. Thus, the *A B C 6th Edition* stated that it was "built on the principle of at least a two-letter difference in each five-letter codeword". The compiler also tried to deal with transposition errors, though admittedly imperfectly; by contrast, the 1923 *Acme Commodity and Phrase Code* [48] proudly stated

> This Code consists of one hundred thousand five-letter code ciphers with at least two-letter

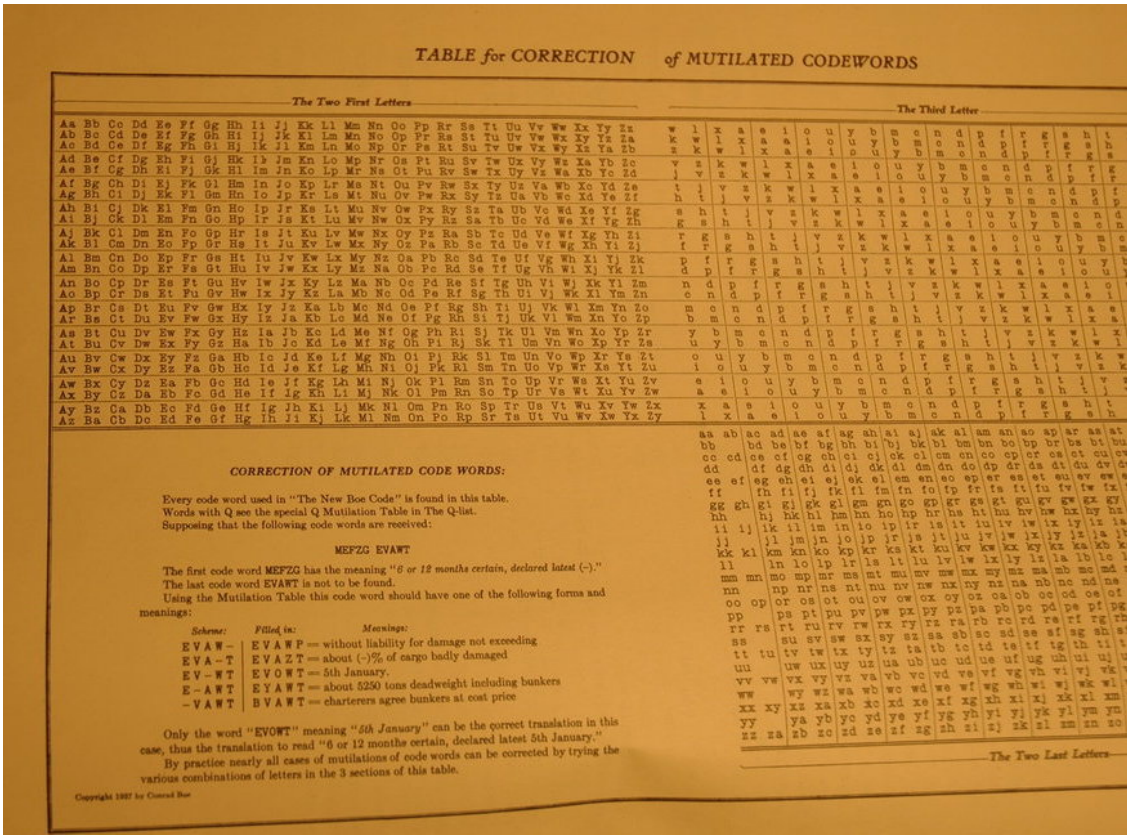


TABLE for CORRECTION of MUTILATED CODEWORDS

The Two First Letters																									The Third Letter										
Aa	Bb	Cc	Dd	Ee	Ff	Gg	Hh	Ii	Jj	Kk	Ll	Mm	Nn	Oo	Pp	Rr	Ss	Tt	Uu	Vv	Ww	Xx	Yy	Zz	w l	x a	e i	o u	y b	m c	n d	p f	r g	s h	t
Ab	Bc	Cd	De	Ef	Fg	Gh	Hi	Ij	Jk	Kl	Lm	Mn	No	Op	Pr	Rs	St	Tu	Uv	Vw	Wx	Xy	Yz	Za	k w	l x	a a	i o	u y	b m	c n	d p	f r	g s	h
Ac	Bd	Ce	Df	Eg	Fh	Gi	Hj	Ik	Jl	Km	Ln	Mo	Np	Or	Ps	Rt	Su	Tv	Uw	Vx	Wy	Xz	Ya	Zb	z k	w l	x a	e i	o u	y b	m c	n d	p f	r g	s
Ad	Be	Cf	Dg	Eh	Fi	Gj	Hk	Il	Jm	Kn	Lo	Mp	Nr	Os	Pt	Ru	Sv	Tw	Ux	Vy	Wz	Xa	Yb	Zc	v z	k w	l x	a e	i o	u y	b m	c n	d p	f r	g
Ae	Bf	Cg	Dh	Ei	Fj	Gk	Hl	Im	Jn	Ko	Lp	Mr	Ns	Ot	Pu	Rv	Sw	Tx	Uy	Vz	Wa	Xb	Yc	Zd	j v	z k	w l	x a	e i	o u	y b	m c	n d	p f	r
Af	Bg	Ch	Di	Ej	Fk	Gl	Hm	In	Jo	Kp	Lr	Ms	Nt	Ou	Pv	Rw	Sx	Ty	Uz	Va	Wb	Xc	Yd	Ze	t j	v z	k w	l x	a e	i o	u y	b m	c n	d p	f
Ag	Bh	Ci	Dj	Ek	Fl	Gm	Hn	Io	Jp	Kr	Ls	Mt	Nu	Ov	Pw	Rx	Sy	Tz	Ua	Vb	Wc	Xd	Ye	Zf	h t	j v	z k	w l	x a	e i	o u	y b	m c	n d	p
Ah	Bi	Cj	Dk	El	Fm	Gn	Ho	Ip	Jr	Ks	Lt	Mu	Nv	Ow	Px	Ry	Sz	Ta	Ub	Vc	Wd	Xe	Yf	Zg	s h	t j	v z	k w	l x	a e	i o	u y	b m	c n	d
Ai	Bj	Ck	Dl	Em	Fn	Go	Hp	Ir	Js	Kt	Lu	Mv	Nw	Ox	Py	Rz	Sa	Tb	Uc	Vd	We	Xf	Yg	Zh	g s	h t	j v	z k	w l	x a	e i	o u	y b	m c	n
Aj	Bk	Cl	Dm	En	Fo	Gp	Hr	Is	Jt	Ku	Lv	Mw	Nx	Oy	Pz	Ra	Sb	Tc	Ud	Ve	Wf	Xg	Yh	Zi	r g	s h	t j	v z	k w	l x	a e	i o	u y	b m	c
Ak	Bl	Cm	Dn	Eo	Fp	Gr	Hs	It	Ju	Kv	Lw	Mx	Ny	Oz	Pa	Rb	Sc	Td	Ue	Vf	Wg	Xh	Yi	Zj	f r	g s	h t	j v	z k	w l	x a	e i	o u	y b	m
Al	Bm	Cn	Do	Ep	Fr	Gs	Ht	Iu	Jv	Kw	Lx	My	Nz	Oa	Pb	Rc	Sd	Te	Uf	Vg	Wh	Xi	Yj	Zk	p f	r g	s h	t j	v z	k w	l x	a e	i o	u y	b
Am	Bn	Co	Dp	Er	Fs	Gt	Hu	Iv	Jw	Kx	Ly	Mz	Na	Ob	Pc	Rd	Se	Tf	Ug	Vh	Wi	Xj	Yk	Zl	d p	f r	g s	h t	j v	z k	w l	x a	e i	o u	y
An	Bo	Cp	Dr	Es	Ft	Gu	Hv	Iw	Jx	Ky	Lz	Ma	Nb	Oc	Pd	Re	Sf	Tg	Uh	Vi	Wj	Xk	Yl	Zm	n d	p f	r g	s h	t j	v z	k w	l x	a e	i o	
Ao	Bp	Cr	Ds	Et	Fu	Gv	Hw	Ix	Jy	Kz	La	Mb	Nc	Od	Pe	Rf	Sg	Th	Ui	Vj	Wk	Xl	Ym	Zn	c n	d p	f r	g s	h t	j v	z k	w l	x a	e i	
Ap	Br	Cs	Dt	Eu	Fv	Gw	Hx	Iy	Jz	Ka	Lb	Mc	Nd	Oe	Pf	Rg	Sh	Ti	Uj	Vk	Wl	Xm	Yn	Zo	m c	n d	p f	r g	s h	t j	v z	k w	l x	a e	
Ar	Bs	Ct	Du	Ev	Fw	Gx	Hy	Iz	Ja	Kb	Lc	Md	Ne	Of	Pg	Rh	Si	Tj	Uk	Vl	Wm	Xn	Yo	Zp	b m	c n	d p	f r	g s	h t	j v	z k	w l	x a	
As	Bt	Cu	Dv	Ew	Fx	Gy	Hz	Ia	Jb	Kc	Ld	Me	Nf	Og	Ph	Ri	Sj	Tk	Ul	Vm	Wn	Xo	Yp	Zr	y b	m c	n d	p f	r g	s h	t j	v z	k w	l x	
At	Bu	Cv	Dw	Ex	Fy	Gz	Ha	Ib	Jc	Kd	Le	Mf	Ng	Oh	Pi	Rj	Sk	Tl	Um	Vn	Wo	Xp	Yr	Zs	u y	b m	c n	d p	f r	g s	h t	j v	z k	w l	
Au	Bv	Cw	Dx	Ey	Fz	Ga	Hb	Ic	Jd	Ke	Lf	Mg	Nh	Oi	Pj	Rk	Sl	Tm	Un	Vo	Wp	Xr	Ys	Zt	o u	y b	m c	n d	p f	r g	s h	t j	v z	k w	
Av	Bw	Cx	Dy	Ez	Fa	Gb	Hc	Id	Je	Kf	Lg	Mh	Ni	Oj	Pk	Rl	Sm	Tn	Uo	Vp	Wr	Xs	Yt	Zu	i o	u y	b m	c n	d p	f r	g s	h t	j v	z k	
Aw	Bx	Cy	Dz	Ea	Fb	Gc	Hd	Ie	Jf	Kg	Lh	Mi	Nj	Ok	Pl	Rm	Sn	To	Up	Vr	Ws	Xt	Yu	Zv	e i	o u	y b	m c	n d	p f	r g	s h	t j	v z	
Ax	By	Cz	Da	Eb	Fc	Gd	He	If	Jg	Kh	Li	Mj	Nk	Ol	Pm	Rn	So	Tp	Ur	Vs	Wt	Xu	Yv	Zw	a e	i o	u y	b m	c n	d p	f r	g s	h t	j	
Ay	Bz	Ca	Db	Ec	Fd	Ge	Hf	Ig	Jh	Ki	Lj	Mk	Nl	Om	Pn	Ro	Sp	Tr	Us	Vt	Wu	Xv	Yw	Zx	x a	e i	o u	y b	m c	n d	p f	r g	s h	t	
Az	Ba	Cb	Dc	Ed	Fe	Gf	Hg	Ih	Ji	Kj	Lk	Ml	Nm	On	Po	Rp	Sr	Ts	Ut	Vu	Wv	Xw	Yx	Zy	l x	a e	i o	u y	b m	c n	d p	f r	g s	h	

CORRECTION OF MUTILATED CODE WORDS:

Every code word used in "The New Boe Code" is found in this table.
Words with Q see the special Q Mutilation Table in The Q-list.
Supposing that the following code words are received:

MEFZG EVAWT

The first code word **MEFZG** has the meaning "*6 or 12 months certain, declared latest* (-)."
The last code word **EVAWT** is not to be found.

Using the Mutilation Table this code word should have one of the following forms and meanings:

Scheme:	*Filled in:*	*Meanings:*
E V A W -	E V A W P	= without liability for damage not exceeding
E V A - T	E V A Z T	= about (-)% of cargo badly damaged
E V - W T	E V O W T	= 5th January.
E - A W T	E Y A W T	= about 5250 tons deadweight including bunkers
- V A W T	B V A W T	= charterers agree bunkers at cost price

Only the word "**EVOWT**" meaning "*5th January*" can be the correct translation in this case, thus the translation to read "6 or 12 months certain, declared latest 5th January."

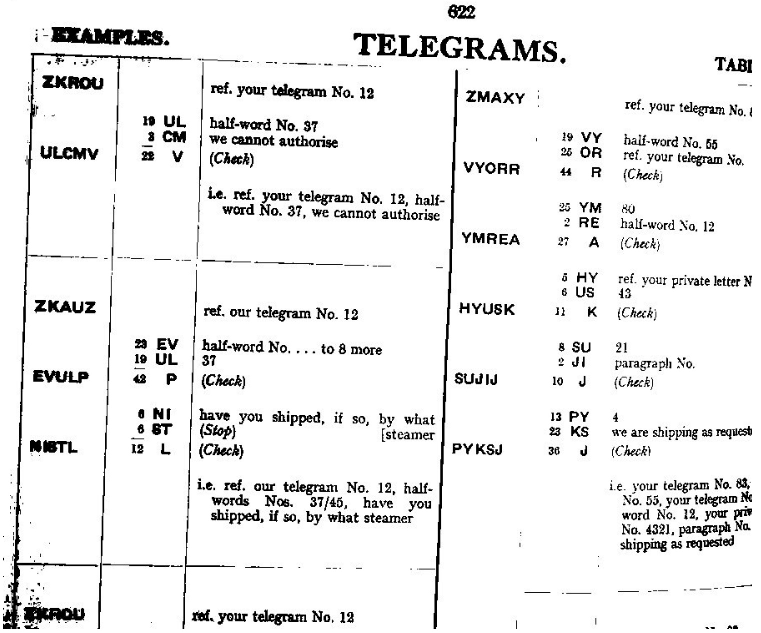
By practice nearly all cases of mutilations of code words can be corrected by trying the various combinations of letters in the 3 sections of this table.

aa ab	ac ad	ae af	ag ah	ai aj	ak al	am an	ao ap	ar as	at
bb	bd be	bf bg	bh bi	bj bk	bl bm	bn bo	bp br	bs bt	bu
cc cd	ce cf	cg ch	ci cj	ck cl	cm cn	co cp	cr cs	ct cu	cv
dd	df dg	dh di	dj dk	dl dm	dn do	dp dr	ds dt	du dv	
ee ef	eg eh	ei ej	ek el	em en	eo ep	er es	et eu	ev ew	
ff	fh fi	fj fk	fl fm	fn fo	fp fr	fs ft	fu fv	fw fx	
gg gh	gi gj	gk gl	gm gn	go gp	gr gs	gt gu	gv gw	gx gy	
hh	hj hk	hl hm	hn ho	hp hr	hs ht	hu hv	hw hx	hy hz	
ii ij	ik il	im in	io ip	ir is	it iu	iv iw	ix iy	iz ia	
jj	jl jm	jn jo	jp jr	js jt	ju jv	jw jx	jy jz	ja jb	
kk kl	km kn	ko kp	kr ks	kt ku	kv kw	kx ky	kz ka	kb	
ll	ln lo	lp lr	ls lt	lu lv	lw lx	ly lz	la lb	lc	
mm mn	mo mp	mr ms	mt mu	mv mw	mx my	mz ma	mb mc	md	
nn	np nr	ns nt	nu nv	nw nx	ny nz	na nb	nc nd	ne	
oo op	or os	ot ou	ov ow	ox oy	oz oa	ob oc	od oe	of	
pp	ps pt	pu pv	pw px	py pz	pa pb	pc pd	pe pf	pg	
rr rs	rt ru	rv rw	rx ry	rz ra	rb rc	rd re	rf rg	rh	
ss	su sv	sw sx	sy sz	sa sb	sc sd	se sf	sg sh		
tt tu	tv tw	tx ty	tz ta	tb tc	td te	tf tg	th ti		
uu	uw ux	uy uz	ua ub	uc ud	ue uf	ug uh	ui uj		
vv vw	vx vy	vz va	vb vc	vd ve	vf vg	vh vi	vj vk		
ww	wy wz	wa wb	wc wd	we wf	wg wh	wi wj	wk wl		
xx xy	xz xa	xb xc	xd xe	xf xg	xh xi	xj xk	xl xm		
yy	ya yb	yc yd	ye yf	yg yh	yi yj	yk yl	ym yn		
zz za	zb zc	zd ze	zf zg	zh zi	zj zk	zl zm	zn zo		

The Two Last Letters

Copyright 1937 by Conrad Boe

Figure 5: A "mutilation table" for error correction.



622

EXAMPLES.

TELEGRAMS.

TABI

ZKROU			ref. your telegram No. 12
	19	UL	half-word No. 37
	3	CM	we cannot authorise
ULCMV	22	V	(*Check*)
			i.e. ref. your telegram No. 12, half-word No. 37, we cannot authorise
ZKAUZ			ref. our telegram No. 12
	23	EV	half-word No. . . . to 8 more
	19	UL	37
EVULP	42	P	(*Check*)
	6	NI	have you shipped, if so, by what
	6	ST	(*Stop*) [steamer
NISTL	12	L	(*Check*)
			i.e. ref. our telegram No. 12, half-words Nos. 37/45, have you shipped, if so, by what steamer
ZKROU			ref. your telegram No. 12

ZMAXY			ref. your telegram No.
	19	VY	half-word No. 55
	25	OR	ref. your telegram No.
VYORR	44	R	(*Check*)
	25	YM	80
	2	RE	half-word No. 12
YMREA	27	A	(*Check*)
	5	HY	ref. your private letter N
	6	US	43
HYUSK	11	K	(*Check*)
	8	SU	21
	2	JI	paragraph No.
SUJIJ	10	J	(*Check*)
	13	PY	4
	23	KS	we are shipping as request
PYKSJ	36	J	(*Check*)
			i.e. your telegram No. 83, No. 55, your telegram No word No. 12, your priv No. 4321, paragraph No. shipping as requested

Figure 6: Check digits for numerical data, from the 1929 *New Standard Code.*

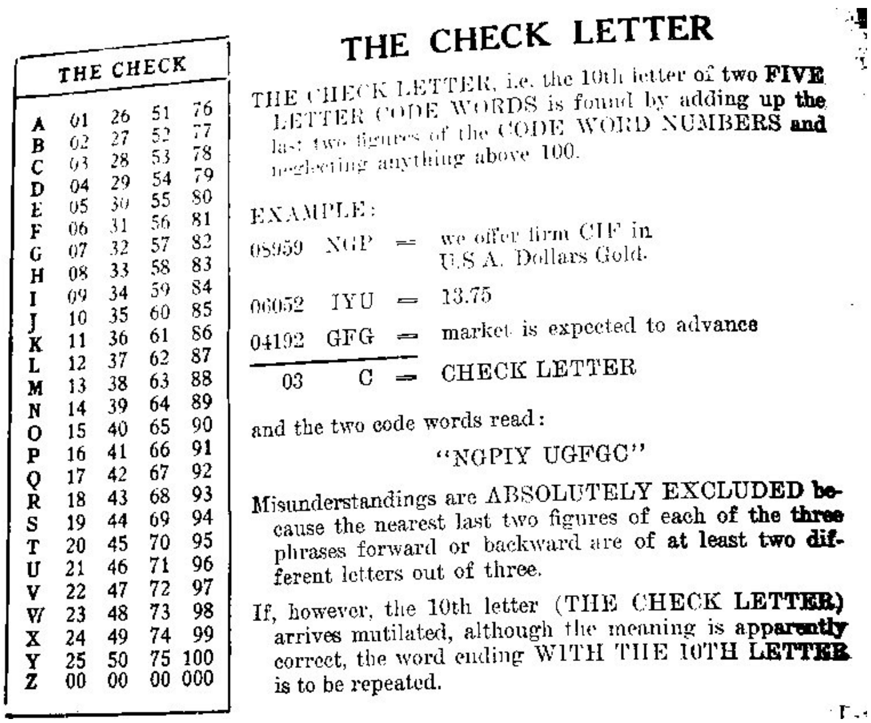
HALAN	HAALN
IBLAN	IBALN
LELAN	LEALN
OGLAN	OGALN
QILAN	QIALN
UMLAN	UMALN
WOLAN	WOALN
ATLAN	ATALN
BULAN	BUALN
EXLAN	EXALN
FYLAN	FYALN

(a) Acme codebook pairs that do not protect against transposition errors.

BEBPY	BEEPY
CIBPY	CIEPY
DOBPY	DOEPY
FUBPY	FUEPY
GABPY	GAEPY
TAUMY	TAZMY
WIUMY	WIZMY
YOUMY	YOZMY

(b) Bentley codebook pairs that are only one character apart.

Figure 7: Codebook compilers did not always succeed in their error-detecting goals. Here we see codeword pairs from the Acme codebook that are not immune to transposition error and pairs from Bentley's code [33] that are only a single letter apart. (Data supplied by J. Reeds.)



THE CHECK				
A	01	26	51	76
B	02	27	52	77
C	03	28	53	78
D	04	29	54	79
E	05	30	55	80
F	06	31	56	81
G	07	32	57	82
H	08	33	58	83
I	09	34	59	84
J	10	35	60	85
K	11	36	61	86
L	12	37	62	87
M	13	38	63	88
N	14	39	64	89
O	15	40	65	90
P	16	41	66	91
Q	17	42	67	92
R	18	43	68	93
S	19	44	69	94
T	20	45	70	95
U	21	46	71	96
V	22	47	72	97
W	23	48	73	98
X	24	49	74	99
Y	25	50	75	100
Z	00	00	00	000

THE CHECK LETTER

THE CHECK LETTER, i.e. the 10th letter of two **FIVE** LETTER CODE WORDS is found by adding **up the** last two figures of the CODE WORD NUMBERS **and** neglecting anything above 100.

EXAMPLE:

08959	NGP	=	we offer firm CIF in U.S.A. Dollars Gold.
06052	IYU	=	13.75
04192	GFG	=	market is expected to advance
03	C	=	CHECK LETTER

and the two code words read:

"NGPIY UGFGC"

Misunderstandings are ABSOLUTELY **EXCLUDED because** the nearest last two figures of each **of the three** phrases forward or backward are of **at least two different** letters out of three.

If, however, the 10th letter (THE CHECK **LETTER**) arrives mutilated, although the meaning is **apparently** correct, the word ending WITH THE 10TH **LETTER** is to be repeated.

Figure 8: Codeword checksum calculation instructions, 1936 *Cosmos Trading Code*.

> difference between each and every word. No transposition of any two adjoining letters will make another word in the book, and we assert that it is the first time this feat has been fully accomplished for 100,000 words.

They did not quite succeed; see Figure 7 for some errors in it and in *Bentley's Complete Phrasebook* [33]. Today, of course, we lump things like the two-letter differential into the general class of Hamming distance [7]; codebooks, though, used the concept several decades before it was formalized.

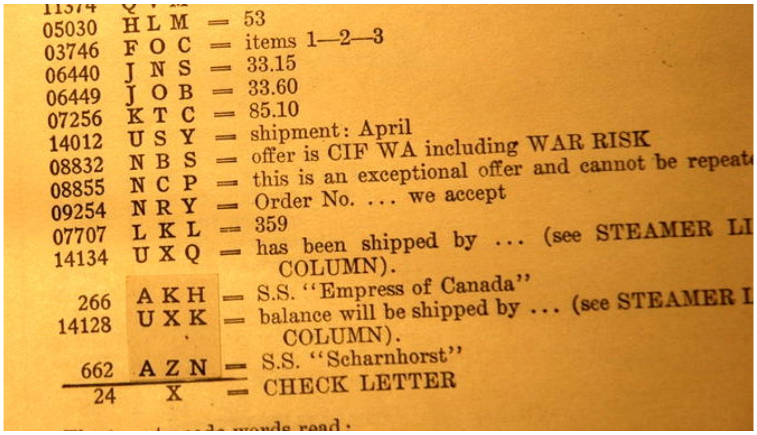
It is worth noting that the check characters used were not as effective against transposition errors as one might want: they generally operated on a mod 10 basis. On the other hand, as a rule two digits were generally encoded at a time, thus providing some protection. Further protection could be gained by ensuring that the second letter of the encoded digraphs could never be the first letter of a numeric code.

That said, hand-calculation of check characters was itself error-prone. Consider the complex process outlined in the *Cosmos Trading Code*, which used a "Three-Letter System". Groups of three letters were combined, with a check letter, into a single telegraph word. But the example supplied in the codebook — Figure 9 — appears to have a correction glued over the last three code entries. Imagine the error rate in production use!

Unencoded numbers were especially subject to corruption during transmission. The 1931 *Swift & Company Private Telegraphic Code* [31], after setting a requirement that telegrams of ten words or more should be coded, says

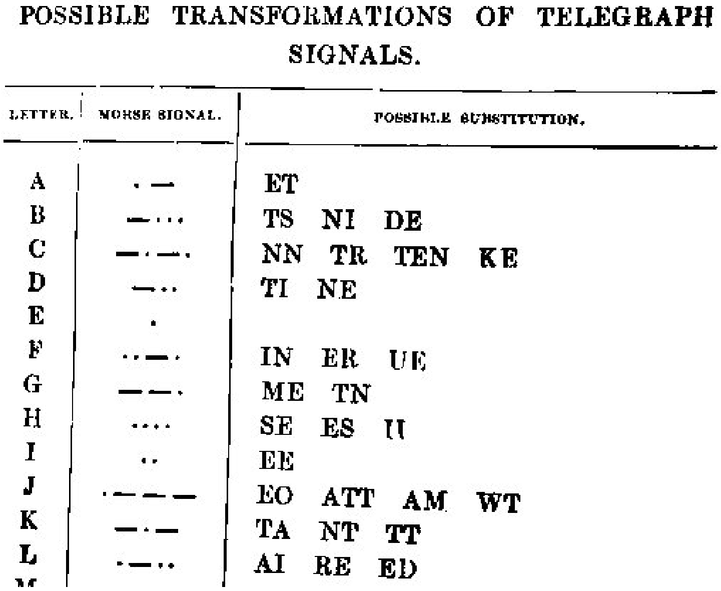
> As a protection against mutilation, phrases, numbers, etc., should be coded if possible, even though the message contains ten words or less. This applies especially to prices and amounts.

Transposition errors are more likely in typed text; in the context of telegrams, this meant when teletypes were used, rather than Morse code. Morse code had its own distinctive errors. Not only could a single dot or dash be omitted, timing variations during transmission could result in a single letter being received as



05030	H L M	=	53
03746	F O C	=	items 1—2—3
06440	J N S	=	33.15
06449	J O B	=	33.60
07256	K T C	=	85.10
14012	U S Y	=	shipment: April
08832	N B S	=	offer is CIF WA including WAR RISK
08855	N C P	=	this is an exceptional offer and cannot be repeat
09254	N R Y	=	Order No. . . . we accept
07707	L K L	=	359
14134	U X Q	=	has been shipped by . . . (see STEAMER LI COLUMN).
266	A K H	=	S.S. "Empress of Canada"
14128	U X K	=	balance will be shipped by . . . (see STEAMER L COLUMN).
662	A Z N	=	S.S. "Scharnhorst"
24	X	=	CHECK LETTER

Figure 9: Correction of errors in the codebook. 1936 *Cosmos Trading Code.*



POSSIBLE TRANSFORMATIONS OF TELEGRAPH SIGNALS.

LETTER.	MORSE SIGNAL.	POSSIBLE SUBSTITUTION.
A	· —	ET
B	— · · ·	TS NI DE
C	— · — ·	NN TR TEN KE
D	— · ·	TI NE
E	·	
F	· · — ·	IN ER UE
G	— — ·	ME TN
H	· · · ·	SE ES II
I	· ·	EE
J	· — — —	EO ATT AM WT
K	— · —	TA NT TT
L	· — · ·	AI RE ED

Figure 10: A table of likely Morse code errors, from the 1886 *Unicode* book.

two. Thus, Figure 10 shows how *F*, transmitted as .._. could be received as *IN* (.. _.), *ER* (. .._.), or *UR* (.._ .).

All of this could be exacerbated by some users' habits of only partially encoding a message. One such instance reached the U.S. Supreme Court, in *Primrose v. Western Union Tel. Co.*, 154 U.S. 1 (1894). Primrose sent the message

> **DESPOT AM EXCEEDINGLY BUSY BAY ALL KINDS QUO PERHAPS BRACKEN HALF OF IT MINCE MOMENT PROMPTLY OF PURCHASES**

Three errors occurred during transmission. "Despot" was received as "Destroy", "bay" was received as "buy", and "purchases" was received in the singular. The second error — a single dot in transmission — was crucial.

Primrose's message was partially encoded. **DESPOT** meant *Yours of the 15th received*; **DESTROY** meant *Yours of the 17th received.* That error was inconsequential. But **BAY** was a codeword which meant *I have bought*; the recipient interpreted **BUY** as the plaintext instruction to purchase some more — and the remainder of the message indicated that 500,00 pounds of wool was the desired quantity.... Primrose lost $20,000 and sued; he lost because of the disclaimer on the back of the telegraph form. Perhaps he could have sued the code compiler — but in fact, it was a private code he and his agent had devised. (More details may be found in [9] or in the Court's opinion.)

The odd appearance of encoded telegrams has had amusing consequences. Once, a New York brokerage firm received an unsigned radiogram reading **ONE LEOPARD AND SEVENTY MONKEYS PERMIT OTHO**. Attempts to decode it using a variety of codebooks failed. The cashier was concerned that a vital trade would be missed, because the market was closing, so he circulated it among the staff. One person finally understood it as plaintext: his son was arriving from Africa on the steamship *Otho* and wanted assistance getting an import permit for a leopard and a large number of monkeys [11].

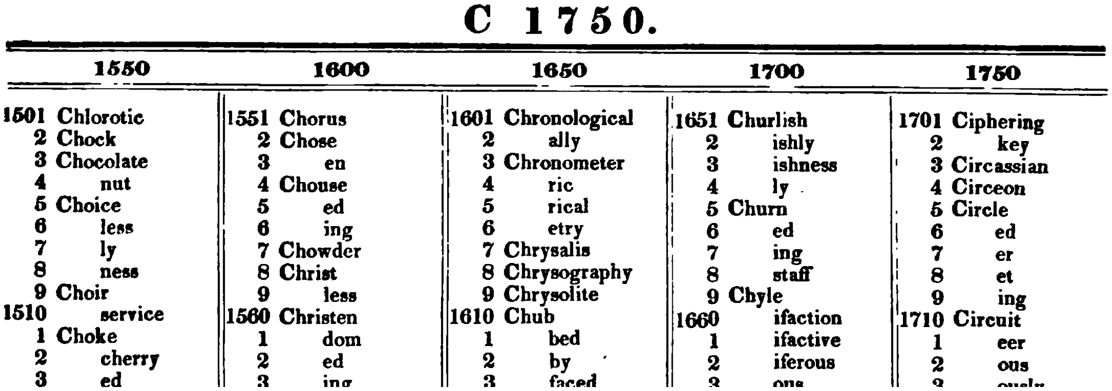
Imposing patterns on codewords, such as a minimum two-letter difference, obviously reduces the size of the space available to compilers. The inclusion of a minimum vowel density requirement reduced it further. An analysis of the actual effect of these constraints was done by Friedman (yes, that Friedman) and Mendelsohn [5].

High-end code makers were aware of such problems, and responded by avoiding use of common words (and especially common commercial words) in their tables. Thus, the 1901 *A B C Telegraphic Code, 5th Edition, Improved* has codewords like **MAELSTROM** and **THEORY**, but not **PURCHASE** or **SELL**. (Not everyone was as careful. The 1900 *Tourists' Telegraphic Code* includes such codewords as **SUBWAY**, **REVOLT**, and **SAVAGERY**. Perhaps well-bred tourists did not encounter the plaintext equivalents!)

It is instructive to consider these problems in the light of modern technology and terminology. The usual sequence of operations today is compression (either generic or domain-specific), encryption, checksum or MAC on the ciphertext, and medium-specific encoding. Each operation is done separately, by a different component, though in some high-performance cryptosystems encryption and MACcing are done in a single pass. In addition, there is generally a checksum at various points during transmission, such as the TCP checksum [16] or the Ethernet CRC. Sound design would suggest an application-level checksum on the plaintext [18]; this is rarely done in most systems.

In telegraph codes, compression was the primary step. The encoding was an integral part of the compression process; more precisely, a separate encoding step was composed with compression, to avoid an extra, expensive, and error-prone pass over the data. Furthermore, the encodings were chosen with particular transmission characteristics in mind. This is not unreasonable — checksums need to be tuned to the medium [21] — but it required changes in encoding (and hence in codebooks) when transmission characteristics changed.

As today, confidentiality was implemented via a transform on the compressed text. This posed a problem, though: since the compression output was already optimally encoded for transmission, a modern-style cipher or even early mechanized encryptors (Enigma, the Hagelin machine, etc.) would have destroyed these properties. Accordingly, the confidentiality systems of the era (see the following section) were effectively a mapping from the codeword space to the codeword space. Today, we seek indistinguishability of a cipher's output from a uniformly distributed random bit string; for telegraph codes, the proper comparison would be a uniformly random selection of codewords.



C 1750.

1550	1600	1650	1700	1750
1501 Chlorotic	1551 Chorus	1601 Chronological	1651 Churlish	1701 Ciphering
2 Chock	2 Chose	2 ally	2 ishly	2 key
3 Chocolate	3 en	3 Chronometer	3 ishness	3 Circassian
4 nut	4 Chouse	4 ric	4 ly	4 Circeon
5 Choice	5 ed	5 rical	5 Churn	5 Circle
6 less	6 ing	6 etry	6 ed	6 ed
7 ly	7 Chowder	7 Chrysalis	7 ing	7 er
8 ness	8 Christ	8 Chrysography	8 staff	8 et
9 Choir	9 less	9 Chrysolite	9 Chyle	9 ing
1510 service	1560 Christen	1610 Chub	1660 ifaction	1710 Circuit
1 Choke	1 dom	1 bed	1 ifactive	1 eer
2 cherry	2 ed	2 by	2 iferous	2 ous
3 ed	3 ing	3 faced	3 ous	3 ously

Figure 11: Part of a page from the *Secret Corresponding Vocabulary* (1845). The compiler, Francis O.J. Smith, was Samuel Morse's business partner in the first commercial deployments of the telegraph. (Image taken from the Google Books digitization of the work.)

4 Confidentiality

Although compression was the primary goal of commercial telegraph codes, confidentiality was a concern as well. Although, as Kahn has noted, the opacity of an ordinary code book was ofen sufficient, many telegraph users required more. In fact, the very first telegraph codebook, Smith's *Secret Corresponding Vocabulary* [50] — though it mentions cost-savings, was intended for confidentiality:

> As the tariff of expense chargeable to correspondents, who shall have recourse to the Telegraph, in order to be equal, can only be based upon the quantity of matter communicated, and as that can only be measured by the number of words transmitted, it is obvious that, in a system where *signs* are employed to represent the letters which form words, whatever will tend to lessen the requisite number of those signs to communicate any given number of words, will add to the despatch of the correspondence, and indirectly, at least, cheapen its transmission.
>
> But, SECRECY in correspondence, is far the most important consideration to be secured. And the crowning desideratum, in the use of the Telegraph, consists in its adaption to this end, by means of the compilation now presented.

The book (Figure 11) contained a list of about 56,000 words. The user would denote a word by its first letter and the index of the word in that section; thus, *ciphering* would be sent as **C.1701**. For confidentiality, a prearranged value was to be added to or subtracted from the index number. Thus, one might send **C.1710** instead if the sender's offset were 9. For more security, a set of different offsets could be used in sequence and a monoalphabetic substitution applied to the letters:

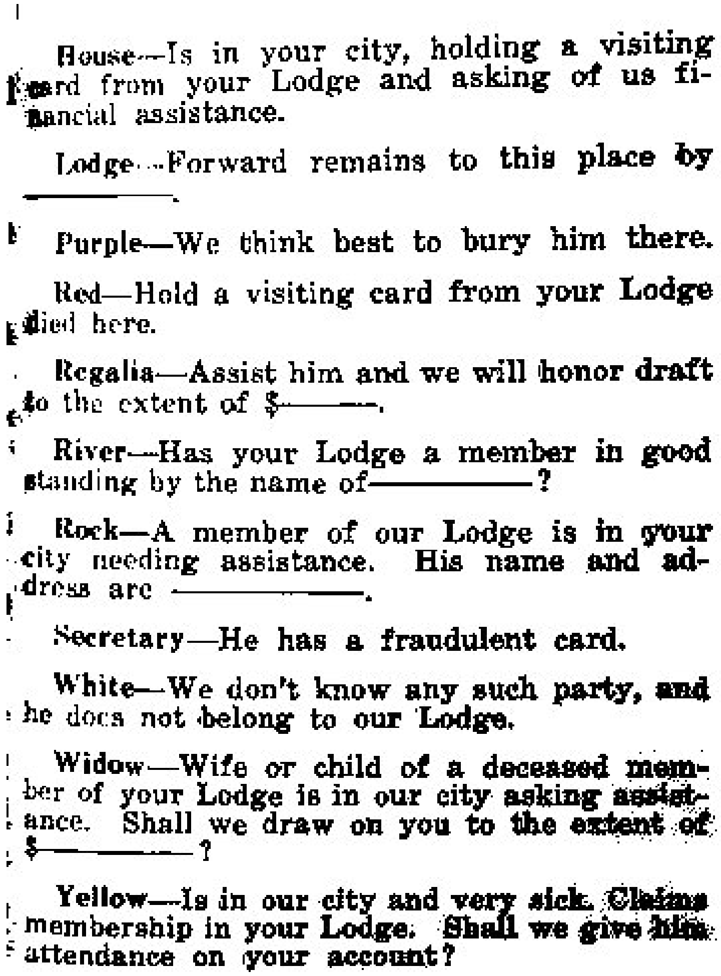
> The complexity of this mode of writing, may be very much increased, so as to render all experiments to decypher communications, utterly hopeless.

Sometimes, though, even for data that might be deemed sensitive, protection against causual readers was deemed sufficient. The 1926 *International Police Telegraph Code* (Figure 4) notes that

> By its nature the Code renders superfluous the translation of an incoming telegram, and so saves valuable police time, while offering a certain guarantee of secrecy.

The recognition of the limitations is itself gratifying.

Needless to say, commercial code confidentiality does not live up to the standards of military or governmental codes. Serious confidentiality codes are "two-part" — separate books or sections are used for



House—Is in your city, holding a visiting card from your Lodge and asking of us financial assistance.

Lodge—Forward remains to this place by ———.

Purple—We think best to bury him there.

Red—Hold a visiting card from your Lodge died here.

Regalia—Assist him and we will honor draft to the extent of $———.

River—Has your Lodge a member in good standing by the name of———?

Rock—A member of our Lodge is in your city needing assistance. His name and address are ———.

Secretary—He has a fraudulent card.

White—We don't know any such party, and he does not belong to our Lodge.

Widow—Wife or child of a deceased member of your Lodge is in our city asking assistance. Shall we draw on you to the extent of $———?

Yellow—Is in our city and very sick. Claims membership in your Lodge. Shall we give him attendance on your account?

Figure 12: A few secrecy code words used by the Independent Order of Odd Fellows, 1931.

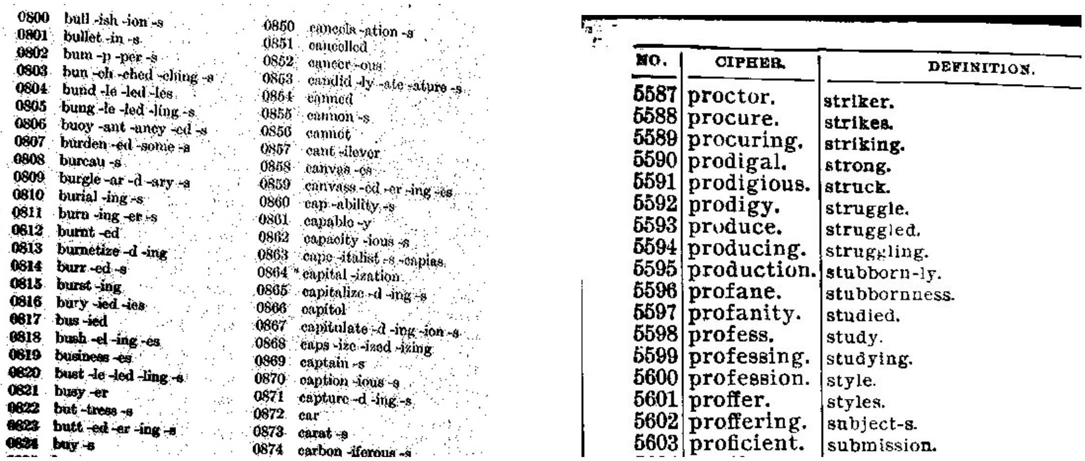
encoding and decoding. This removes the requirement, clearly shown in these examples, that the plaintext and the code words be in the same order. Other measures commonly taken include multiple ciphertext symbols for common plaintext phrases and superencipherment of the codewords. Only the latter was commonly used commercially, and rarely well.

For simple uses, secrecy of the code words was employed. A fraternal group, the Independent Order of Odd Fellows, published a 1931 constitution and bylaws booklet that included two pages of a 1908 "Telegraph Cipher and Key" (Figure 12).

The normal approach was superencipherment. An amusing pair is the New York Central Lines' 1923 *Van Code*, which is "to be used only when secrecy is desired" (Figure 13a) and the 1892 *Sheahan's Telegraphic Cipher Code* (Figure 13b)

> "for use of the several Organizations of Railway Employes [sic] . . . 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 . . . as it is often necessary to use the Company's wire.

Of the two, labor employed better technique. The key was an integer added to the code number; the code word corresponding to the new code number was to be sent. Thus, if the key were 3 and someone wanted to



0800	bull -ish -ion -s	0850	cancels -ation -s
0801	bullet -in -s	0851	cancelled
0802	bum -p -per -s	0852	cancer -ous
0803	bun -ch -ched -ching -s	0853	candid -ly -ate -ature -s
0804	bund -le -led -les	0854	canned
0805	bung -le -led -ling -s	0855	cannon -s
0806	buoy -ant -ancy -ed -s	0856	cannot
0807	burden -ed -some -s	0857	cant -ilever
0808	bureau -s	0858	canvas -es
0809	burgle -ar -d -ary -s	0859	canvass -ed -er -ing -es
0810	burial -ing -s	0860	cap -ability -s
0811	burn -ing -er -s	0861	capable -y
0812	burnt -ed	0862	capacity -ious -s
0813	burnetize -d -ing	0863	cape -italist -s -capias
0814	burr -ed -s	0864	capital -ization
0815	burst -ing	0865	capitalize -d -ing -s
0816	bury -ied -ies	0866	capitol
0817	bus -ied	0867	capitulate -d -ing -ion -s
0818	bush -el -ing -es	0868	caps -ize -ized -izing
0819	business -es	0869	captain -s
0820	bust -le -led -ling -s	0870	caption -ious -s
0821	busy -er	0871	capture -d -ing -s
0822	but -tress -s	0872	car
0823	butt -ed -er -ing -s	0873	carat -s
0824	buy -s	0874	carbon -iferous -s

(a)

NO.	CIPHER.	DEFINITION.
5587	proctor.	striker.
5588	procure.	strikes.
5589	procuring.	striking.
5590	prodigal.	strong.
5591	prodigious.	struck.
5592	prodigy.	struggle.
5593	produce.	struggled.
5594	producing.	struggling.
5595	production.	stubborn-ly.
5596	profane.	stubbornness.
5597	profanity.	studied.
5598	profess.	study.
5599	professing.	studying.
5600	profession.	style.
5601	proffer.	styles.
5602	proffering.	subject-s.
5603	proficient.	submission.

(b)

Figure 13: Secrecy codebooks for railroad use, (a) management and (b) labor.

send the word *struggle*, 3 would be added to **5592** and *production* would be transmitted in its place. Users were cautioned never to mix plaintext and ciphertext.

The most intriguing part of the scheme was the deliberate omission of numbers (and hence superencipherment) for times and dates, for fear of known plaintext attacks:

> This plan was adopted after careful study and deliberation, as a safeguard for the reason that a telegram giving a number a name, or in reference to anything that occurrred on a certain day would, if the same key number applied to the entire book, be a clew [sic] that would lead to the discovery of your key number. Therefore, I have used numbers only where I believed it was safe to do so.

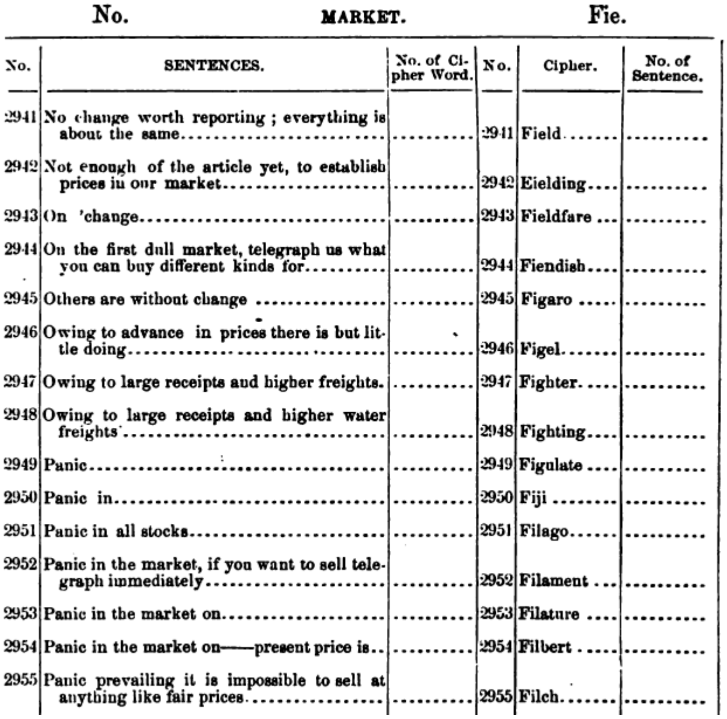
It is unclear how successful this book was, organizationally or cryptographically; however, it was reissued at least as late as 1912, and perhaps 1938.

Management showed much less sophistication. Mixed ciphertext and plaintext was expressly supported, and keying was a choice of either sending "the word opposite" or the "Arbitrary Word to the left" of the desired word.

Bloomer's Commercial Cryptograph: A Telegraph Code and Double Index-holocryptic Cipher (1874) [34] showed more cryptographic sophistication than many. In addition to the usual additives, it suggested transposition of code words. The practical effect of that would often be minimal, especially on short messages — **ABUKIR FILAGO EVACUATE** (*Advice from New York; panic in all stocks; market affected by general causes*) would be nearly as intelligible if rendered as *Market affected by general causes; panic in all stocks; advice from New York.* The scheme would do considerably better if different additives were used for successive words, a technique that is also described.

More interestingly, Bloomer appears to have understood the benefit of two-part codes. The codebook (Figure 14) provided extra spaces for each code word and code sentence, with the following advice:

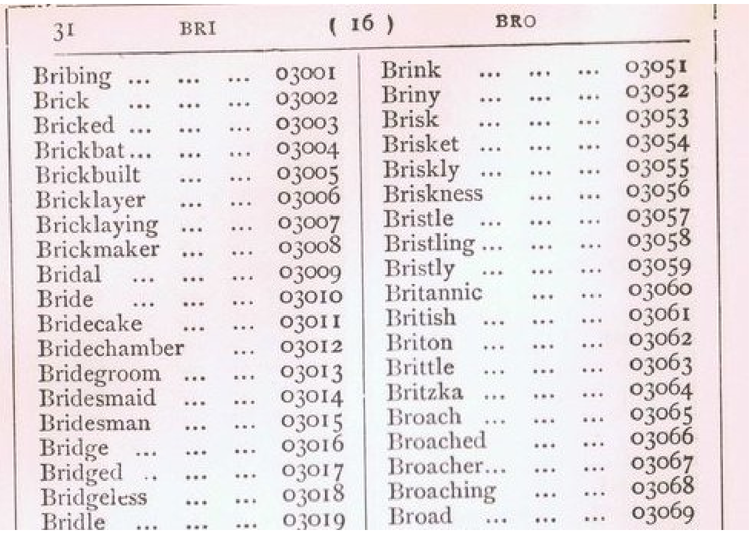
> 5th.—Double Index—A permanent cryptograph may he made in the third and sixth columns by selecting cipher words indiscriminately from the fifth column, and entering the numbers of such words in the third column, opposite the sentences which the cipher words are intended to represent, and entering the numbers of such sentences in the sixth column, opposite the cipher words selected. Thus, if 2228 be written in the third column, opposite numher 2175, "Buy at seller's option," and 2175 in the sixth column, opposite the cipher word "Doctor," the party



No. MARKET. Fie.

No.	SENTENCES.	No. of Cipher Word.	No.	Cipher.	No. of Sentence.
2941	No change worth reporting ; everything is about the same..........................	....	2941	Field......	....
2942	Not enough of the article yet, to establish prices iu onr market..................	....	2942	Eielding....	....
2943	On 'change...............................	....	2943	Fieldfare ...	....
2944	On the first dull market, telegraph us what you can buy different kinds for.........	....	2944	Fiendish....	....
2945	Others are without change	....	2945	Figaro	....
2946	Owing to advance in prices there is but little doing.............................	....	2946	Figel.......	....
2947	Owing to large receipts aud higher freights.	....	2947	Fighter.....	....
2948	Owing to large receipts and higher water freights..............................	....	2948	Fighting....	....
2949	Panic..................................	....	2949	Figulate	....
2950	Panic in................................	....	2950	Fiji	....
2951	Panic in all stocks........................	....	2951	Filago......	....
2952	Panic in the market, if you want to sell telegraph immediately......................	....	2952	Filament ...	....
2953	Panic in the market on....................	....	2953	Filature	....
2954	Panic in the market on——present price is..	....	2954	Filbert	....
2955	Panic prevailing it is impossible to sell at anything like fair prices.................	....	2955	Filch.......	....

Figure 14: An excerpt from Bloomer's Commercial Cryptograph. Note the blank spaces for writing in variant code numbers.



31 BRI (16) BRO

Bribing	03001	Brink	03051
Brick	03002	Briny	03052
Bricked	03003	Brisk	03053
Brickbat	03004	Brisket	03054
Brickbuilt	03005	Briskly	03055
Bricklayer	03006	Briskness	03056
Bricklaying	03007	Bristle	03057
Brickmaker	03008	Bristling	03058
Bridal	03009	Bristly	03059
Bride	03010	Britannic	03060
Bridecake	03011	British	03061
Bridechamber	03012	Briton	03062
Bridegroom	03013	Brittle	03063
Bridesmaid	03014	Britzka	03064
Bridesman	03015	Broach	03065
Bridge	03016	Broached	03066
Bridged	03017	Broacher	03067
Bridgeless	03018	Broaching	03068
Bridle	03019	Broad	03069

Figure 15: Code numbers for Slater's secrecy code.

> desiring to telegraph the above sentence will find the numher 2228. By turning to the printed number 2228, the cipher word will he found to be "Doctor," which being telegraphed, the receiver finds opposite "Doctor," 2175, the number of the original sentence. In this case it will he necessary for correspondents to have the exact copy of the numbers written in both volumes.

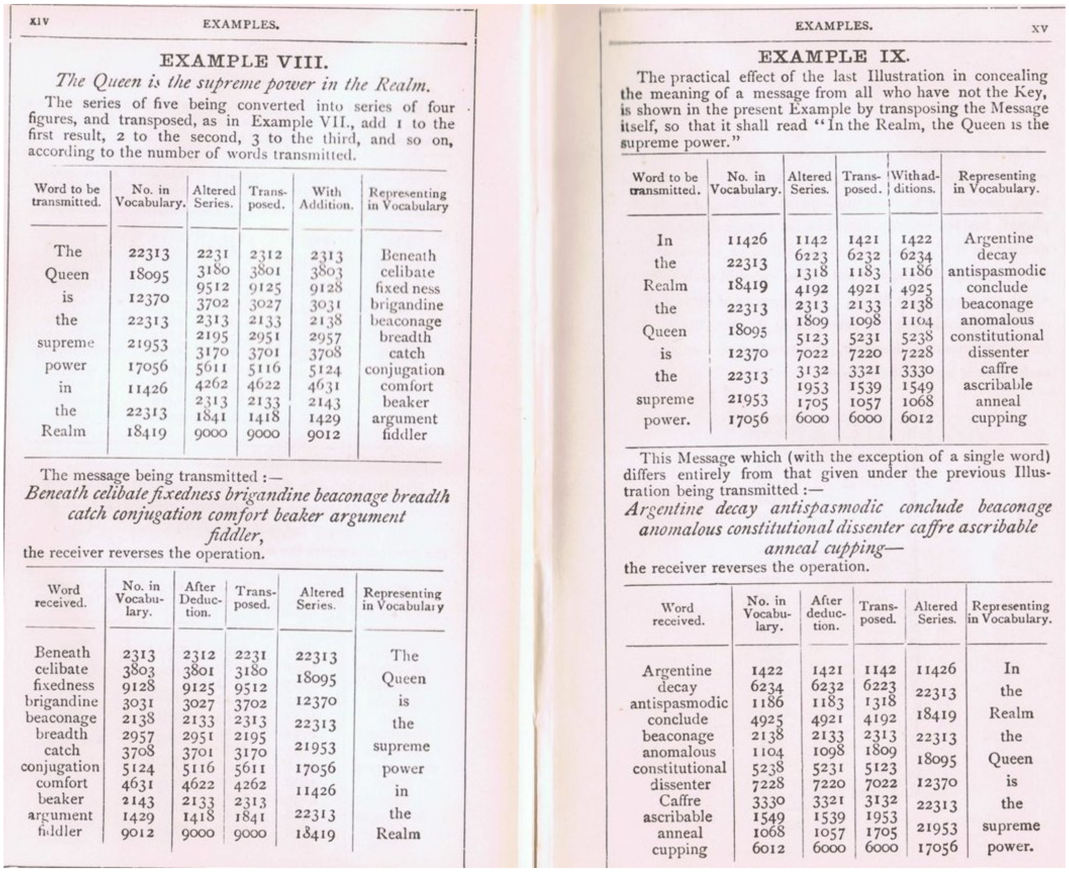
In other words, users of the code were instructed to create their own two-part equivalences, pair by pair. This is a laborious process, and of dubious utility unless many such pairs were created. It may safely be assumed that very few users actually carried out this process to any noticeable extent.

Most of the commercial codebooks offer add-ons that promise "absolute secrecy". These tend to be simple transforms of the codeword or code number, or monoalphabetic transformations of the individual characters of the code word. There is one, though, that stands out: *Slater's Telegraphic Code*, since it was intended solely for secrecy and provided no compression or correction. Conventional wisdom has it that there was never a market for commercial secrecy; this codebook, though, lasted from about 1870 until at least 1938. The threat model was interesting as well:

> 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. Now in large or small communities (particularly perhaps in the latter) there are always to be found prying spirits, curious as to the affairs of their neighbours, which they think they can manage so much better than the parties chiefly interested, and proverbially inclined to gossip.

To start, a message was converted to code numbers via the book (Figure 15). Next, some transform was applied to the sequence of code numbers. Several types are suggested: simple addition or subtraction of a key number, transposition of some of the ciphertext digits, and regrouping into four-character sections instead of five. Combinations also suggested. Of particular interest is the realization by the compiler that with regrouping, minor changes in plaintext can result in very large changes of ciphertext (Figure 16).

It is tempting to laugh at the cryptanalytic naivete of these code compilers. At least in the U.S., the military did no better back then. In 1899, the War Department published a supplement to the Western



XIV EXAMPLES.

EXAMPLE VIII.

The Queen is the supreme power in the Realm.

The series of five being converted into series of four figures, and transposed, as in Example VII., add 1 to the first result, 2 to the second, 3 to the third, and so on, according to the number of words transmitted.

Word to be transmitted.	No. in Vocabulary.	Altered Series.	Transposed.	With Addition.	Representing in Vocabulary
The	22313	2231	2312	2313	Beneath
Queen	18095	3180	3801	3803	celibate
		9512	9125	9128	fixed ness
is	12370	3702	3027	3031	brigandine
the	22313	2313	2133	2138	beaconage
supreme	21953	2195	2951	2957	breadth
		3170	3701	3708	catch
power	17056	5611	5116	5124	conjugation
in	11426	4262	4622	4631	comfort
		2313	2133	2143	beaker
the	22313	1841	1418	1429	argument
Realm	18419	9000	9000	9012	fiddler

The message being transmitted :—

Beneath celibate fixedness brigandine beaconage breadth catch conjugation comfort beaker argument fiddler,

the receiver reverses the operation.

Word received.	No. in Vocabulary.	After Deduction.	Transposed.	Altered Series.	Representing in Vocabulary
Beneath	2313	2312	2231	22313	The
celibate	3803	3801	3180	18095	Queen
fixedness	9128	9125	9512		
brigandine	3031	3027	3702	12370	is
beaconage	2138	2133	2313	22313	the
breadth	2957	2951	2195		
catch	3708	3701	3170	21953	supreme
conjugation	5124	5116	5611	17056	power
comfort	4631	4622	4262	11426	in
beaker	2143	2133	2313		
argument	1429	1418	1841	22313	the
fiddler	9012	9000	9000	18419	Realm

EXAMPLES. XV

EXAMPLE IX.

The practical effect of the last Illustration in concealing the meaning of a message from all who have not the Key, is shown in the present Example by transposing the Message itself, so that it shall read "In the Realm, the Queen is the supreme power."

Word to be transmitted.	No. in Vocabulary.	Altered Series.	Transposed.	With additions.	Representing in Vocabulary.
In	11426	1142	1421	1422	Argentine
the	22313	6223	6232	6234	decay
		1318	1183	1186	antispasmodic
Realm	18419	4192	4921	4925	conclude
the	22313	2313	2133	2138	beaconage
		1809	1098	1104	anomalous
Queen	18095	5123	5231	5238	constitutional
is	12370	7022	7220	7228	dissenter
the	22313	3132	3321	3330	caffre
		1953	1539	1549	ascribable
supreme	21953	1705	1057	1068	anneal
power.	17056	6000	6000	6012	cupping

This Message which (with the exception of a single word) differs entirely from that given under the previous Illustration being transmitted :—

Argentine decay antispasmodic conclude beaconage anomalous constitutional dissenter caffre ascribable anneal cupping—

the receiver reverses the operation.

Word received.	No. in Vocabulary.	After deduction.	Transposed.	Altered Series.	Representing in Vocabulary.
Argentine	1422	1421	1142	11426	In
decay	6234	6232	6223	22313	the
antispasmodic	1186	1183	1318		
conclude	4925	4921	4192	18419	Realm
beaconage	2138	2133	2313	22313	the
anomalous	1104	1098	1809	18095	Queen
constitutional	5238	5231	5123		
dissenter	7228	7220	7022	12370	is
Caffre	3330	3321	3132	22313	the
ascribable	1549	1539	1953		
anneal	1068	1057	1705	21953	supreme
cupping	6012	6000	6000	17056	power.

Figure 16: Some suggested transformations of code numbers.

Union code book [30] until their own full code could be compiled [29] five years later. Although economy was the primary concern, “it is also to be used as a cipher code in important and confidential messages where secrecy is desired” [45]. The suggested scheme? “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.” The codeword corresponding to the new number was then used. Kahn calls this “probably the most secure and advanced code system of the day” [9, p. 252].

The 1899 U.S. Navy [13] felt the same way about mixing plaintext and ciphertext:

> In order to eliminate as far as possible errors in transmission due to mistakes of telegraphic operators in telegraphing words strange to them, it is hereby directed that in using the cipher code only that part of the communication which is of a confidential nature be put in cipher, except in cases where the cipher code is used to shorten the message in order that the telegraphic cost may be materially lessened.

Arguably, the State Department was even worse. They were much more concerned with economy than confidentiality [24, 25], and their codes reflected that. Superencryption schemes, similar to Bloom’s, were provided as an appendix; given diplomats’ penchant for sending mixed plaintext and codewords, one can assume that these schemes were seldom used. Not surprisingly, other countries were frequently able to read U.S. diplomatic traffic. Indeed, in a note that Roosevelt sent to Japan pleading for peace on December 6, 1941, he specified that a known-insecure code be used because

> . . . he did not mind if the dispatch was “picked up”, and also that the code “saves time”. [24].

When Roosevelt wanted security, he had the Navy transmit his messages [9].

Even addresses were considered sensitive sometimes, though no solution was propounded. Companies could register short addresses with their telegraph companies, much as domain names are used today. New York, unlike many cities, had a central list serving all companies. They had had separate lists, but “in 1917, the State Department, fearing spies, abolished all existing lists and set up a uniform one for everybody” [2].

Full names were important as well. In the Bahamas during World War II, people were required to sign their full names on telegrams, even those going to family members [17].

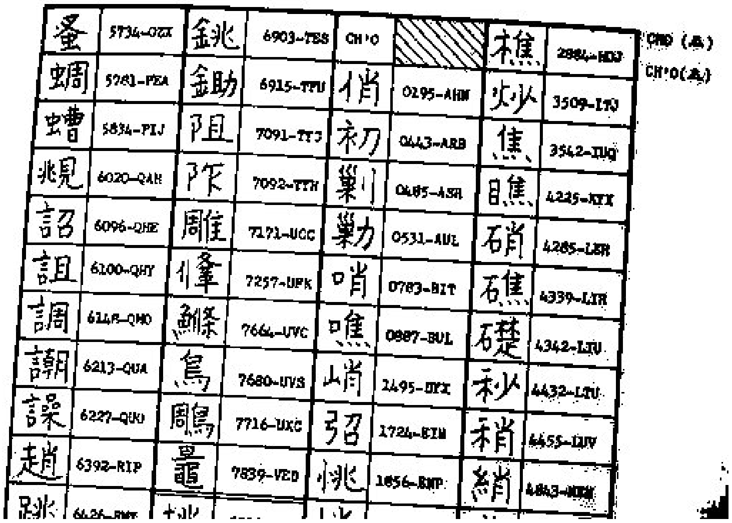
5 Comprehension

This title of this section refers many forms of comprehension. Under this heading I’ve lumped linguistic issues, coding issues, and — most important — what we learn of other cultures, removed in time from ours.

The simplest issue was character set suitability. Any alphabetic script, whether Latin, Greek, Cyrillic, or Hebrew, can be transmitted rather easily. Ideographic languages, such as Chinese, Japanese, and Korean, pose serious issues for the telegraph operator. The primary purpose of such a codebook (Figure 17) is simply encoding into an alphabetic form, often on a per-word basis. On top of that, phrase compression and substitution could be added (Figure 18).

The need for such encodings has not vanished. Telegrams were popular in China until about 10 years ago. The codebook used — typically, a 4-digit encoding was employed for each ideograph — was originally developed in the 1870s. Although telegraph usage has dropped off sharply in recent years, as mobile phones have become extremely common, the codebooks are still used for spelling names in certain circumstances, such as when applying for a passport. Many Chinese characters are very similar-looking; this form of encoding is less ambiguous and dialect-independent [28]. Its use is often recommended for police use, to avoid errors from transliteration: a name’s sound (and hence its transliteration) are dialect-dependent; ideographs are not [3]. The same set of code points are used for other purposes, such as machine translation. However, the evolution of the Chinese language over time — new words, and hence new ideographs, and the switch to simplified ideographs that the government of the People’s Republic of China started in the mid-1950s — has resulted in a more complex code [27, 1], with newer entries not in their nominally-correct place.

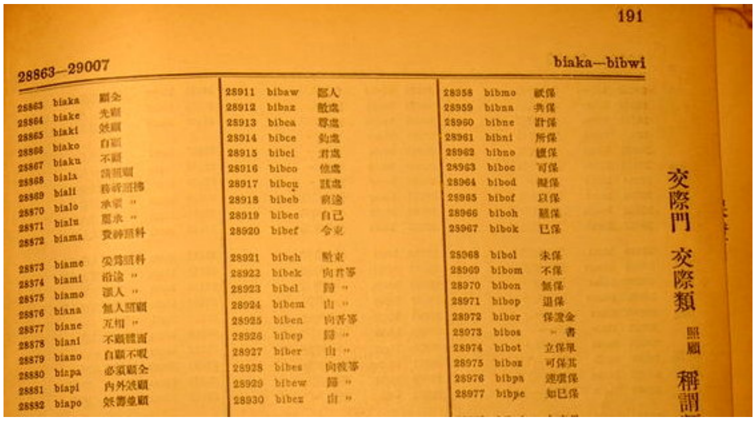
More sophisticated codes were multilingual, where the codeword provided the mapping between languages. Thus, in the 1923 *Peterson International Code, 2nd Edition* (Figure 19) [49], **FYOUG** is rendered



蚤	5734-OZX	銚	6903-TES	CH'O	▨	樵	2984-HDJ
蜩	5781-PEA	鋤	6915-TFU	俏	0295-AHN	炒	3509-ITQ
螬	5834-PIJ	阻	7091-TTJ	初	0443-ARB	焦	3542-IUQ
覜	6020-QAH	阼	7092-TTH	剿	0485-ASR	瞧	4225-KYX
詔	6096-QHE	雕	7171-UCC	勦	0531-AUL	硝	4285-LER
詛	6100-QHY	鞗	7257-UFK	哨	0783-BIT	礁	4339-LIR
調	6148-QMO	鰷	7664-UVC	噍	0887-BUL	礎	4342-LIU
謿	6213-QUA	鳥	7680-UVS	峭	1495-DYX	秒	4432-LTU
譟	6227-QUU	鵰	7716-UXC	弨	1724-EIN	稍	4455-LUV
趙	6392-RIP	鼂	7839-VED	恌	1856-ENP	綃	4843-NIK
跳	6426-RMT						

CHO (조)
CH'O(초)

Figure 17: The *Korean Telegraphic Code.* The book isn't dated; however, it says it uses the McCune-Reischauer romanization system, which was first published in 1939 [12]. The Library of Congress catalog lists 1950, 1952, and 1978 books of that name.

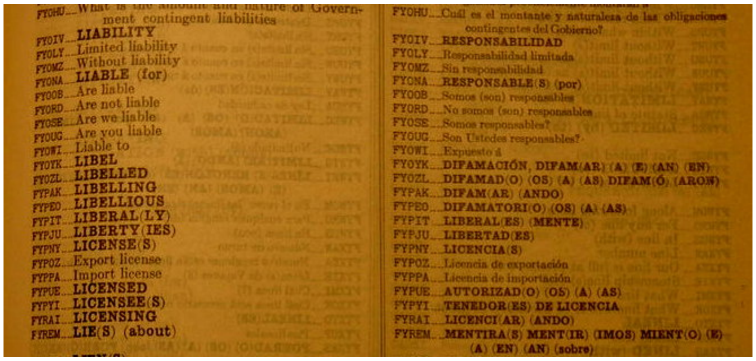


28863—29007 191 biaka—bibwi

28863	biaka	顧全	28911	bibaw	託人	28958	bibmo	祗保
28864	biake	先顧	28912	bibaz	散處	28959	bibna	共保
28865	biaki	兼顧	28913	bibea	尊處	28960	bibne	討保
28866	biako	自顧	28914	bibce	鈞處	28961	bibni	所保
28867	biaku	不顧	28915	bibei	君處	28962	bibno	擔保
28868	biala	請照顧	28916	bibeo	他處	28963	biboc	可保
28869	biali	務祈照拂	28917	bibeu	該處	28964	bibod	擬保
28870	bialo	承蒙 〃	28918	bibeb	前途	28965	bibof	具保
28871	bialu	懇求 〃	28919	bibec	自己	28966	biboh	願保
28872	biama	費神照料	28920	bibef	令東	28967	bibok	已保
28873	biame	受累照料	28921	bibeh	賠東	28968	bibol	未保
28874	biami	沿途 〃	28922	bibek	向君等	28969	bibom	不保
28875	biamo	該人 〃	28923	bibel	歸 〃	28970	bibon	無保
28876	biana	無人照顧	28924	bibem	由 〃	28971	bibop	退保
28877	biane	互相 〃	28925	biben	向吾等	28972	bibor	保證金
28878	biani	不顧體面	28926	bibep	歸 〃	28973	bibos	〃 書
28879	biano	自顧不暇	28927	biber	由 〃	28974	bibot	立保單
28880	biapa	必須顧全	28928	bibes	向彼等	28975	biboz	可保其
28881	biapi	內外兼顧	28929	bibew	歸 〃	28976	bibpa	速覓保
28882	biapo	兼籌並顧	28930	bibez	由 〃	28977	bibpe	如已保

交際門 交際類 照顧 稱謂

Figure 18: The 1915 *China Republican Telegraphic Code.*



FYOHU	What is the amount and nature of Government contingent liabilities	FYOHU	Cuál es el montante y naturaleza de las obligaciones contingentes del Gobierno?
FYOIV	**LIABILITY**	FYOIV	**RESPONSABILIDAD**
FYOLY	Limited liability	FYOLY	Responsabilidad limitada
FYOMZ	Without liability	FYOMZ	Sin responsabilidad
FYONA	**LIABLE (for)**	FYONA	**RESPONSABLE(S) (por)**
FYOOB	Are liable	FYOOB	Somos (son) responsables
FYORD	Are not liable	FYORD	No somos (son) responsables
FYOSE	Are we liable	FYOSE	Somos responsables?
FYOUG	Are you liable	FYOUG	Son Ustedes responsables?
FYOWI	Liable to	FYOWI	Expuesto á
FYOYK	**LIBEL**	FYOYK	**DIFAMACIÓN, DIFAM(AR) (A) (E) (AN) (EN)**
FYOZL	**LIBELLED**	FYOZL	**DIFAMAD(O) (OS) (A) (AS) DIFAM(Ó) (ARON)**
FYPAK	**LIBELLING**	FYPAK	**DIFAM(AR) (ANDO)**
FYPEO	**LIBELLIOUS**	FYPEO	**DIFAMATORI(O) (OS) (A) (AS)**
FYPIT	**LIBERAL(LY)**	FYPIT	**LIBERAL(ES) (MENTE)**
FYPJU	**LIBERTY(IES)**	FYPJU	**LIBERTAD(ES)**
FYPNY	**LICENSE(S)**	FYPNY	**LICENCIA(S)**
FYPOZ	Export license	FYPOZ	Licencia de exportación
FYPPA	Import license	FYPPA	Licencia de importación
FYPUE	**LICENSED**	FYPUE	**AUTORIZAD(O) (OS) (A) (AS)**
FYPYI	**LICENSEE(S)**	FYPYI	**TENEDOR(ES) DE LICENCIA**
FYRAI	**LICENSING**	FYRAI	**LICENCI(AR) (ANDO)**
FYREM	**LIE(S) (about)**	FYREM	**MENTIRA(S) MENT(IR) (IMOS) MIENT(O) (E) (A) (EN) (AN) (sobre)**

Figure 19: The 1923 *Peterson International Code, 2nd Edition* with both English and Spanish plaintexts [49].

as both *Are you liable* and *Son Ustedes responsables?*. Kahn describes codebooks that encompassed nine languages.

Note the difficulty that one-part codes have when multilingual: the plaintext alphabetization cannot be consistent. Word indices were provided, often for all languages, to help with that problem.

Kahn notes that "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." Consider the many forms of household staff members described in the 1926 *International Police Telegraph Code* (Figure 20). How many of those entries were relevant even five years later, given onset of the Great Depression?

Specialized professions had their own codes. The theater world had one [40]; it included many variants on phrases like **DISORB**, which meant *do not want drunkards.* Some phrases would probably not appear in a modern equivalent, such as **DORIAN** for *Jew comedian.* **FILIATION**, *chorus girls who are shapely and good looking*, might appear today, though I suspect that **FILIBUSTER** — *chorus girls who are shapely, good looking, and can sing* might displace it. Then and now, though, the large section on "Financial Straits" is probably appropriate.

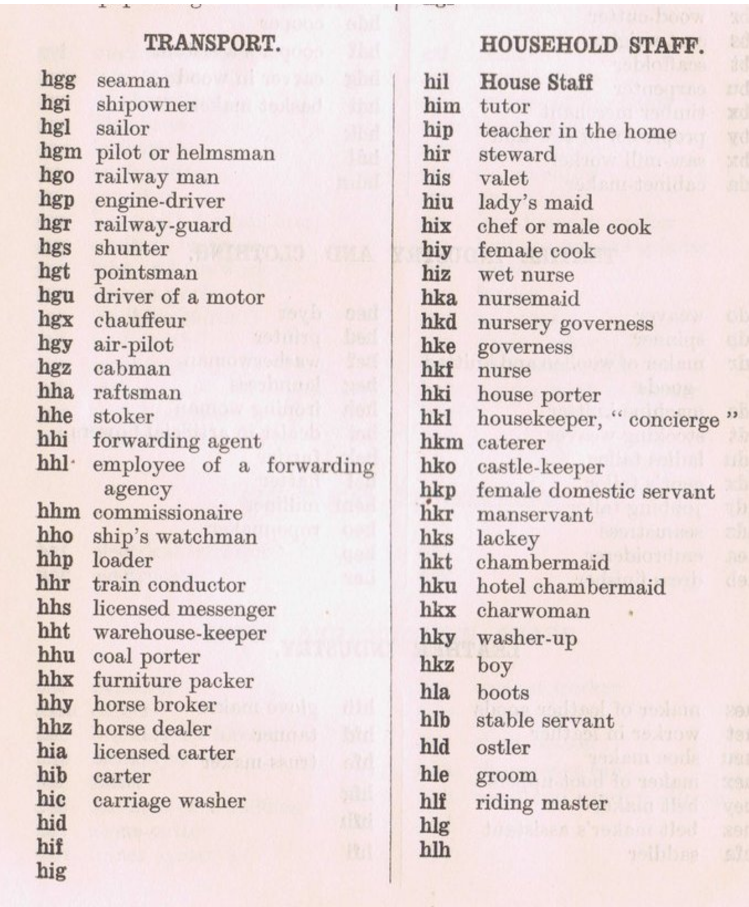
The old naval codes are just as illuminating. The section for "Person's Names" in [51], for example, includes *Earl of * *.* No other titles of nobility except *Queen* are listed; presumably, *Lord* and *Lady* are expected to suffice. Popham [44] has a signal for *Send women on shore to wash.*

A more light-hearted example may be seen in Figure 21. How often, today, does one speak of a marriage being arranged between two parties? (Of course, that is a culture-centric statement, too; in many parts of the world, arranged marriages are still the norm.) Surprisingly modern concepts can show up; the 1926 police code did include *living together with* in the same grouping as marriages.

The same Unicode book suggests that telegrams were a very rapid means of communication: there is a code phrase (Figure 23) used for scheduling a lunch that same day. Think of the steps necesssary: the message must be composed, encoded, delivered to a telegraph office for transmission, delivered to the recipient, and decoded. Email is much simpler!

The most fascinating code, from a perspective of revealing attitudes, is the *China Inland Mission Private Telegraph Code* (1907) [36, 37, 38]. It is replete with many references to "natives" (**19316** is *The natives in the district are very troublesome*), many phrases about the addressee's wife but none about husbands, etc. It was a turbulent era in China; the Boxer Rebellion had just ended and the Revolution of 1912 was about to start. Not surprisingly, there are many phrases concerning disturbances, riots, rebellions, and revolutions. The most fascinating phrases, though, are **23697** and **23699**, about Catholics (Figure 22).

The necessity of using the telegraph was reflected throughout society. Indeed, there are those who argue that relatively speaking, the telegraph had far more effect in its day than the Internet has had today [20]. Today, catalogs frequently contain URLs. In 1936, the Norton Company's *List Prices of Norton Grinding*



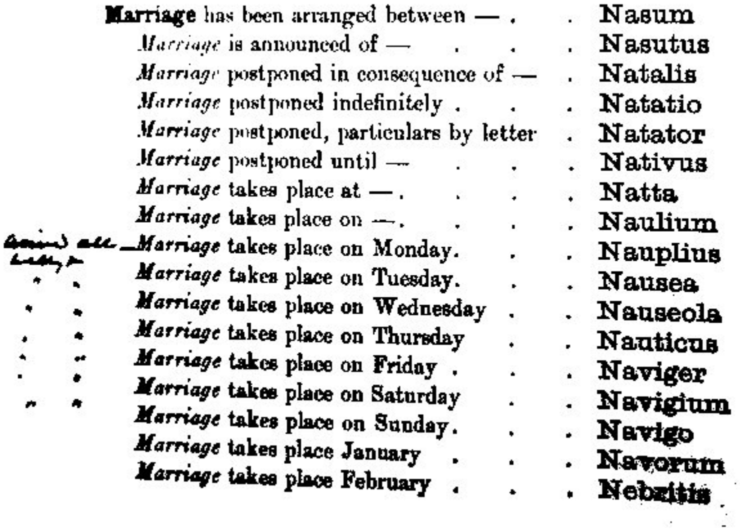
TRANSPORT.

hgg	seaman
hgi	shipowner
hgl	sailor
hgm	pilot or helmsman
hgo	railway man
hgp	engine-driver
hgr	railway-guard
hgs	shunter
hgt	pointsman
hgu	driver of a motor
hgx	chauffeur
hgy	air-pilot
hgz	cabman
hha	raftsman
hhe	stoker
hhi	forwarding agent
hhl	employee of a forwarding agency
hhm	commissionaire
hho	ship's watchman
hhp	loader
hhr	train conductor
hhs	licensed messenger
hht	warehouse-keeper
hhu	coal porter
hhx	furniture packer
hhy	horse broker
hhz	horse dealer
hia	licensed carter
hib	carter
hic	carriage washer
hid	
hif	
hig	

HOUSEHOLD STAFF.

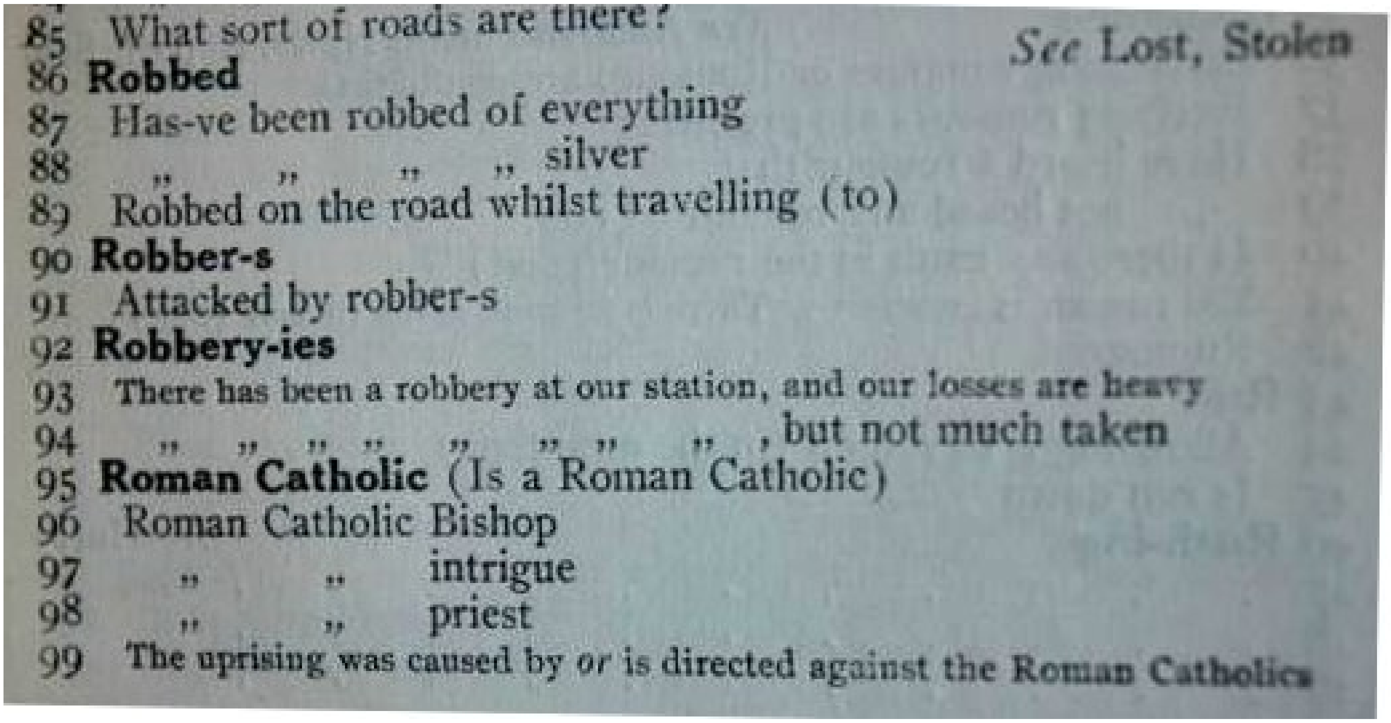
hil	**House Staff**
him	tutor
hip	teacher in the home
hir	steward
his	valet
hiu	lady's maid
hix	chef or male cook
hiy	female cook
hiz	wet nurse
hka	nursemaid
hkd	nursery governess
hke	governess
hkf	nurse
hki	house porter
hkl	housekeeper, " concierge "
hkm	caterer
hko	castle-keeper
hkp	female domestic servant
hkr	manservant
hks	lackey
hkt	chambermaid
hku	hotel chambermaid
hkx	charwoman
hky	washer-up
hkz	boy
hla	boots
hlb	stable servant
hld	ostler
hle	groom
hlf	riding master
hlg	
hlh	

Figure 20: Some professions in the *International Police Telegraph Code*.



Marriage has been arranged between — . .	**Nasum**
Marriage is announced of — . . .	**Nasutus**
Marriage postponed in consequence of — .	**Natalis**
Marriage postponed indefinitely . . .	**Natatio**
Marriage postponed, particulars by letter .	**Natator**
Marriage postponed until — . . .	**Nativus**
Marriage takes place at —	**Natta**
Marriage takes place on —	**Naulium**
[illegible] all [illegible] *Marriage* takes place on Monday. . .	**Nauplius**
Marriage takes place on Tuesday. . .	**Nausea**
Marriage takes place on Wednesday . .	**Nauseola**
Marriage takes place on Thursday . .	**Nauticus**
Marriage takes place on Friday . . .	**Naviger**
Marriage takes place on Saturday . .	**Navigium**
Marriage takes place on Sunday. . .	**Navigo**
Marriage takes place January . . .	**Navorum**
Marriage takes place February . . .	**Nebritis**

Figure 21: How common were arranged marriages in 1886 (*Unicode*)?



85 What sort of roads are there?

See Lost, Stolen

86 **Robbed**

87 Has-ye been robbed of everything

88 „ „ „ „ silver

89 Robbed on the road whilst travelling (to)

90 **Robber-s**

91 Attacked by robber-s

92 **Robbery-ies**

93 There has been a robbery at our station, and our losses are heavy

94 „ „ „ „ „ „ „ „ , but not much taken

95 **Roman Catholic** (Is a Roman Catholic)

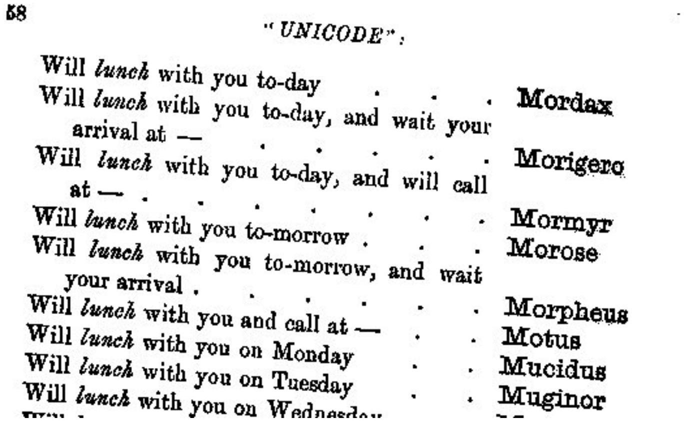
96 Roman Catholic Bishop

97 „ „ intrigue

98 „ „ priest

99 The uprising was caused by *or* is directed against the Roman Catholics

Figure 22: The code of the China Inland Mission — a Protestant group — shows some attitudes towards the Catholic “competition”. (Image from [37].)

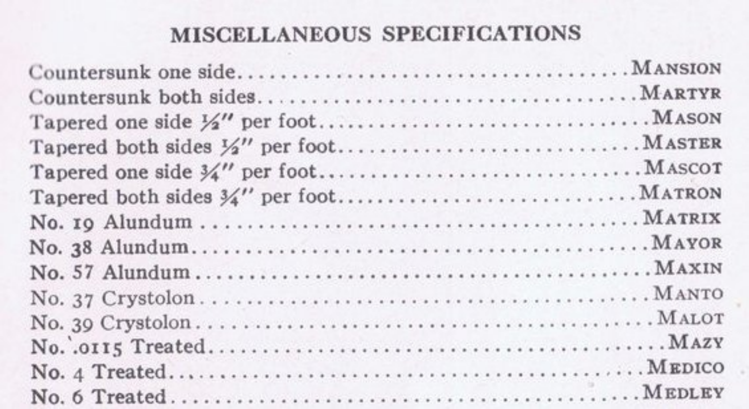


58

“UNICODE”:

Will *lunch* with you to-day	**Mordax**
Will *lunch* with you to-day, and wait your arrival at —	**Morigero**
Will *lunch* with you to-day, and will call at —	**Mormyr**
Will *lunch* with you to-morrow	**Morose**
Will *lunch* with you to-morrow, and wait your arrival	**Morpheus**
Will *lunch* with you and call at —	**Motus**
Will *lunch* with you on Monday	**Mucidus**
Will *lunch* with you on Tuesday	**Muginor**
Will *lunch* with you on Wednesd	

Figure 23: Telegrams really were encoded, sent, received, and decoded in a single morning in 1886 (*Unicode*).



MISCELLANEOUS SPECIFICATIONS

Countersunk one side	Mansion
Countersunk both sides	Martyr
Tapered one side ½" per foot	Mason
Tapered both sides ½" per foot	Master
Tapered one side ¾" per foot	Mascot
Tapered both sides ¾" per foot	Matron
No. 19 Alundum	Matrix
No. 38 Alundum	Mayor
No. 57 Alundum	Maxin
No. 37 Crystolon	Manto
No. 39 Crystolon	Malot
No. .0115 Treated	Mazy
No. 4 Treated	Medico
No. 6 Treated	Medley

Figure 24: Norton grinding wheels.

Wheels included code words (Figure 24) for each product.

Codebooks were expensive. The 1896 *Atlas Universal Travelers' and Business Telegraphic Cipher Code* sold for $5.00 only a few years before New York's Biltmore Hotel, in an ad on the outside cover of the *A B C Telegraphic Code, 5th Edition* [39] offered rooms starting at $2.50. (The ad also noted that 950 of their 1000 rooms had baths. Perhaps the $2.50 rooms were in the lower 5%...)

The cost of many of the codebooks was indeed defrayed by advertising. Figure 25 shows one in the 1920 *A B C 6th Edition* [42]. Most of the ads were, as to be expected, aimed at businesses or at least business travelers, but — as shown here — there were exceptions.

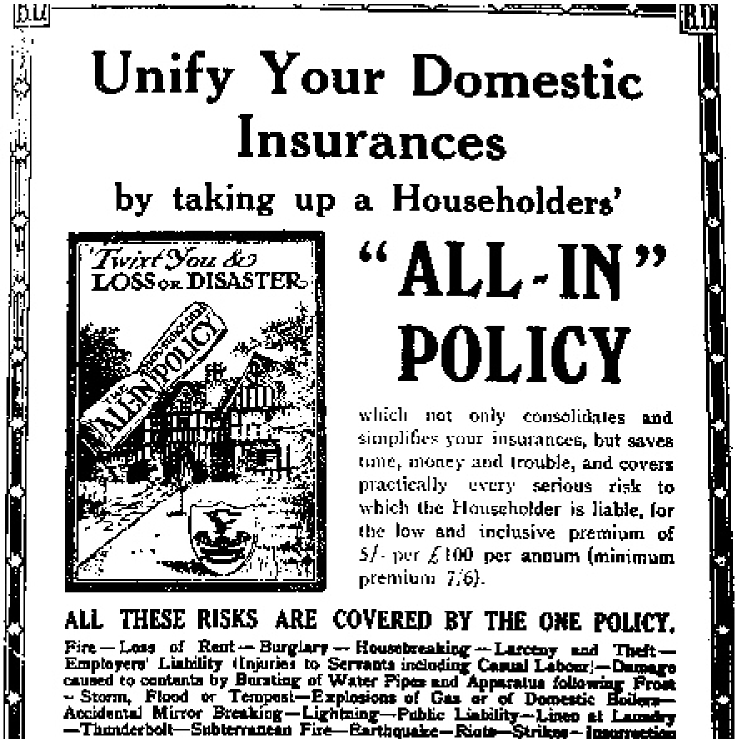
In an eerie parallel with today's controversies, communications intelligence gleaned from telegraphed messages was quite important. Kahn tells much of this story, but the geopolitical aspects are even more fascinating. The U.K. was the center of the world's cable lines; a very high percentage of international messages flowed through Britain or or one of its colonies [8]. This was by design. Not only were "All-red routes" — so-called because that was the map color used for the British Empire — preferred to protect domestic traffic, the Official Secrets Act of 1920 required cable companies to turn over to the government copies of all international telegrams. One U.S. executive tried to explain away the problem in some Senate hearings:

> The messages were then placed in large bags, sealed I believe, and put in wagons. These wagons were driven away under the custody of the Admiralty and lodged overnight in a storehouse and returned to the cable offices the next morning. So that they were kept — they had actual custody of the messages but for a few hours, and so far as the United States messages were concerned, only as a matter of form to make the custom uniform for all countires. We have further investigated and are satisfied that during that period not a single message, commercial, diplomatic, or otherwise, has been actually handled by the Naval Intelligence Bureau, and that their contents are unknown to the British Government because of that fact.

[8] goes on to wonder if he was "the most naive person ever to testify before Congress, or the most deceitful".

The military importance of civilian telegraph codes continued. During World War II, the Army's Signals Intelligence Service had a Commercial Code Unit in the Code Recovery Section of the Cryptanalytic Branch [26]. The data they obtained provided insight into economic conditions in various countries, as well as providing trade and travel data.

Finally, we note another important point of commonality with that era: intellectual property rights were a battleground then, too. The 1936 *A B C 7th Edition* [43] offered a reward for information about infringers (Figure 26). Perhaps more significantly, different legal standards in different countries led to problems. For protectionist reasons, U.S. copyright law did not protect books unless they were printed within the country;



Unify Your Domestic Insurances

by taking up a Householders'

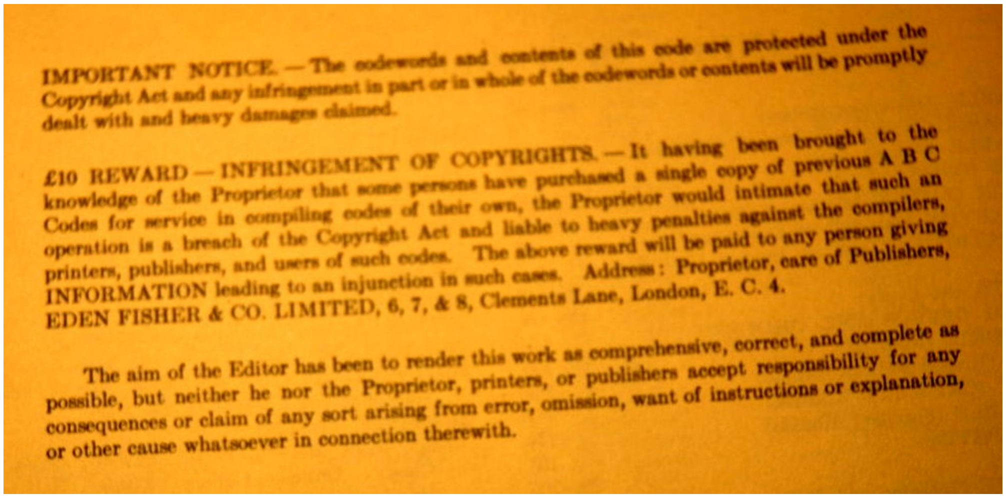
"ALL-IN" POLICY

which not only consolidates and simplifies your insurances, but saves time, money and trouble, and covers practically every serious risk to which the Householder is liable, for the low and inclusive premium of 5/- per £100 per annum (minimum premium 7/6).

ALL THESE RISKS ARE COVERED BY THE ONE POLICY.

Fire — Loss of Rent — Burglary — Housebreaking — Larceny and Theft — Employers' Liability (Injuries to Servants including Casual Labour) — Damage caused to contents by Bursting of Water Pipes and Apparatus following Frost — Storm, Flood or Tempest — Explosions of Gas or of Domestic Boilers — Accidental Mirror Breaking — Lightning — Public Liability — Linen at Laundry — Thunderbolt — Subterranean Fire — Earthquake — Riots — Strikes — Insurrection

Figure 25: An ad in the *A B C 6th Edition* codebook [42].

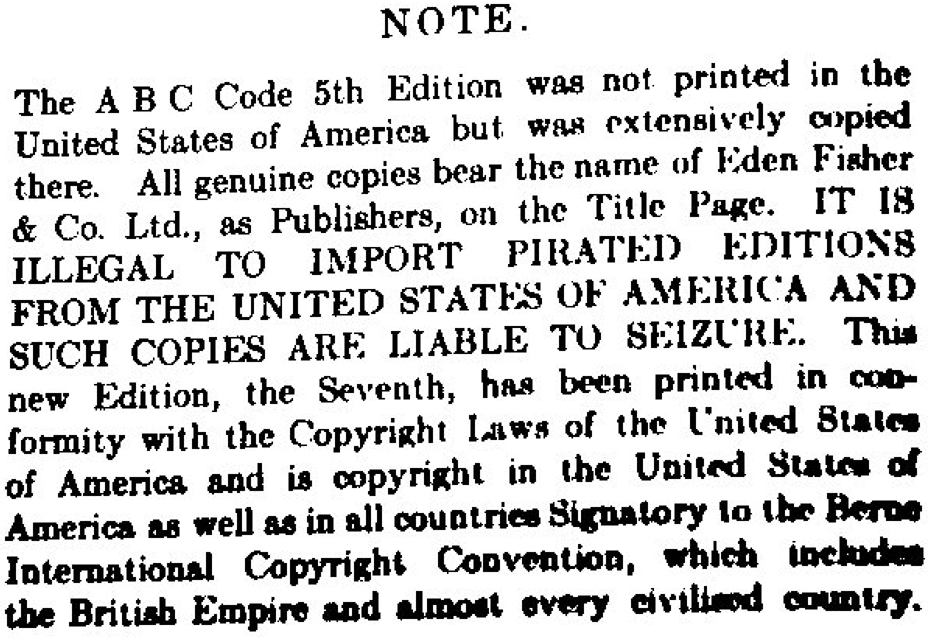


IMPORTANT NOTICE. — The codewords and contents of this code are protected under the Copyright Act and any infringement in part or in whole of the codewords or contents will be promptly dealt with and heavy damages claimed.

£10 REWARD — INFRINGEMENT OF COPYRIGHTS. — It having been brought to the knowledge of the Proprietor that some persons have purchased a single copy of previous A B C Codes for service in compiling codes of their own, the Proprietor would intimate that such an operation is a breach of the Copyright Act and liable to heavy penalties against the compilers, printers, publishers, and users of such codes. The above reward will be paid to any person giving INFORMATION leading to an injunction in such cases. Address: Proprietor, care of Publishers, EDEN FISHER & CO. LIMITED, 6, 7, & 8, Clements Lane, London, E. C. 4.

The aim of the Editor has been to render this work as comprehensive, correct, and complete as possible, but neither he nor the Proprietor, printers, or publishers accept responsibility for any consequences or claim of any sort arising from error, omission, want of instructions or explanation, or other cause whatsoever in connection therewith.

Figure 26: A copyright issue in 1936...



NOTE.

The A B C Code 5th Edition was not printed in the United States of America but was extensively copied there. All genuine copies bear the name of Eden Fisher & Co. Ltd., as Publishers, on the Title Page. IT IS ILLEGAL TO IMPORT PIRATED EDITIONS FROM THE UNITED STATES OF AMERICA AND SUCH COPIES ARE LIABLE TO SEIZURE. This new Edition, the Seventh, has been printed in conformity with the Copyright Laws of the United States of America and is copyright in the United States of America as well as in all countries Signatory to the Berne International Copyright Convention, which includes the British Empire and almost every civilised country.

Figure 27: Was the U.S. civilised then?

this could and did lead to piracy. The wording of the warning in Figure 27 suggests that perhaps this was not a civilised [sic] country then…

6 The Transmission Stack

Figure 28 shows the system architecture as a network stack. Three of our concepts — confidentiality, compression, and correction — can be applied at any of the layers.

At first blush, this seems odd; at the plaintext layer, there would appear to be little room for any of it. Indeed, we now realize that for information-theoretic reasons, we cannot compress encrypted text; applying any sort of confidentiality transform before using the code books would seem impossible. Perhaps more to the point, given the semantic nature of the codes, compression would seem unlikely as well. Still, it can be done, by operating at the semantic level. In *McNeill's Code, 1908 Edition* [47], users desiring secrecy are told to combine certain numeric fields based on semantic knowledge. Thus, a day of the month — two digits — and the maximum daily output of a stamping mill (asserted to be three digits at most — can be combined into a single five-digit number, for which there is a code word equivalent. This reduces the number of groups by one; it also makes life harder for an enemy cryptanalyst.

The vocabulary of codebooks force the user to do other types of compression. Nelson, for example, originally asked that the signal be sent as "England confides that every man will do his duty". Upon being informed that "confides" was not in the codebook but that "expects" was, he agreed to the variant wording.

Finally, the structure of the code may itself force compression. In [35], correspondents are instructed to use a particular stylized form for routine reports of disease outbreak: the date, the port or location, a list of disease-number pairs, a list of ports followed by **UB** to indicate no plague, cholera, smallpox, or yellow fever, and the title of the person filing the report.

[46], rather than being a code book per se, is an algorithm and a set of tables for code construction and use. The user is instructed to compile a list of stylized phrases, perhaps of the form subjects, verbs, and

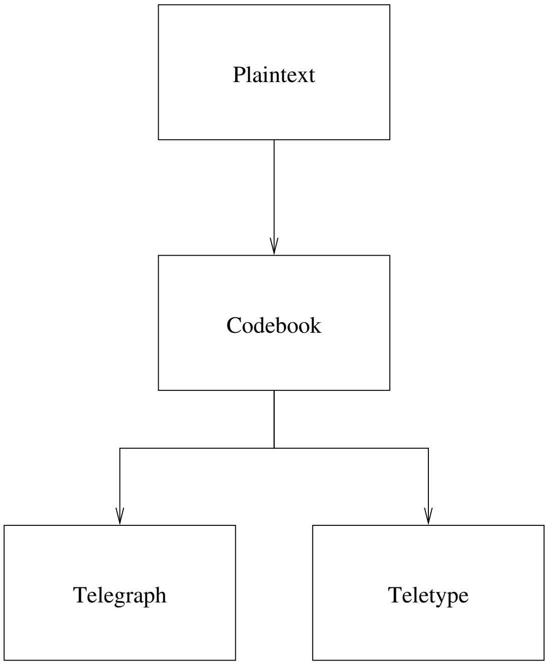


Figure 28: A stack architecture for communications.

objects. Each list entry is denoted by a letter or number; to decode a triplet, the reader would look up each letter in the appropriate list. Many such sets of lists are possible; an indicator word is sent first to denote in what order the sets should be consulted.

Error-handling is more of a stretch; still, one can note the rules given in [31]:

> Care must be exercised in copying words from the Code.
>
> Each code word should begin with a capital letter and should be written distinctly (typewritten, if possible) to prevent errors in transmission.

Perhaps more to the point, code words, especially for numerical quantities were much less error-prone than the actual plaintext. The same book thus instructs

> As a protection against mutilation, phrases, numbers, etc., should be coded if possible, even though the message contains ten words or less. This applies especially to prices and amounts.

It is obvious how all three functions are accomplished via codebooks; we will not belabor the point further. It is, though, worth noting that [10] asserts that well before the middle of the 20th century, advances in cryptanalysis had doomed the use of codebooks for protection against sophisticated enemies. Commercial codebooks, even if enciphered, would offer no protection at all.

During the era of telegraph code books, little specific was done to provide link-level confidentiality, at least for commercial messages. Indeed, confidential diplomatic messages were sent by the same means. Compression and error protection were important, but in a non-obvious way. That said, Friedman notes "certain firms . . . at the present time prefer to use wire and cable telegraphy exclusively [as opposed to radio] and must, for purposes of secrecy, as is the case with banks and brokerage house, use code" [4]. In other words, link selection was done in part to increase confidentiality, because available, economical, technical mechanisms were perceived to be inadequate.

Compression must always be done against some metric. Today, we are concerned with net bits per second, perhaps with a tradeoff against latency or computational power. The primary metric then was cost — what

```
  b     a    n    e    f     u     l
 _...   ._   _.   .   .._.  .._  ._..
--------------------------------------
  d     u    t    i    f     u     l
 _..   .._   _    ..  .._.  .._  ._..
```

Figure 29: Friedman [4] shows how the movement of two dots between adjacent letters can completely change the appearance of a word.

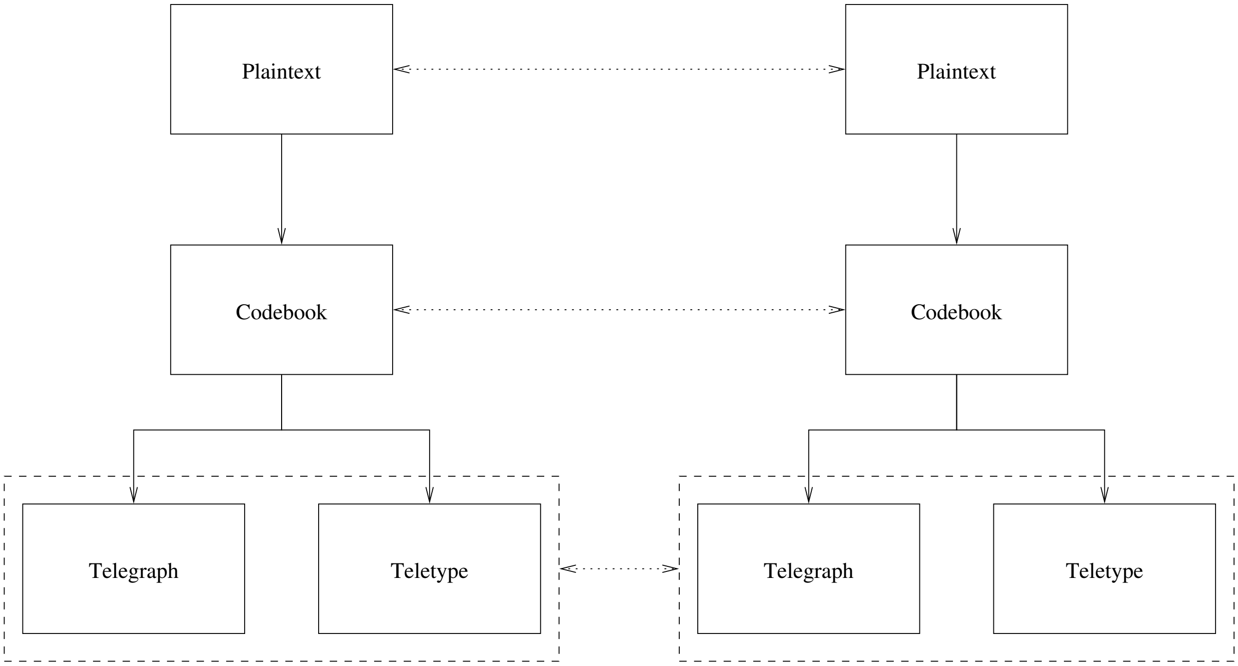


Figure 30: Communication between entities at different levels.

the telegraph companies would charge for a given message — with error rates the primary tradeoff. Usually, the charge was per word; that, however, rests on the definition of “word”. Internationally, at least, this was a matter of treaties and regulations; these changed over time. (See [4] for details of the issues.) When the rules permitted words in any of fifty or so languages, code compilers used many such dictionaries. When the rules permitted words that were “pronounceable” according to the rules of eight different languages, code makers adapted to that. In one example, [35] specified that the letter **A** should be freely inserted into code words to make them pronounceable, and should be deleted on receipt. More commonly, code books were often designed so that two or more code groups could be combined into one chargeable word.

Not surprisingly, error characteristics are heavily affected by link characteristics. Figure 29 shows how the appearance of a word can be completely changed by common Morse code mistakes. Clearly, the switch to teletypes would completely eliminate that sort of error. On the other hand, new types of errors, such as accidentally hitting an adjacent letter on the keyboard, could occur. Error-correcting mechanisms needed to be redesigned accordingly.

Another way to look at the situation is to realize that stack components communicate with their peers. Their properties — or failures — in confidentiality, compression, or correction are first manifested at that

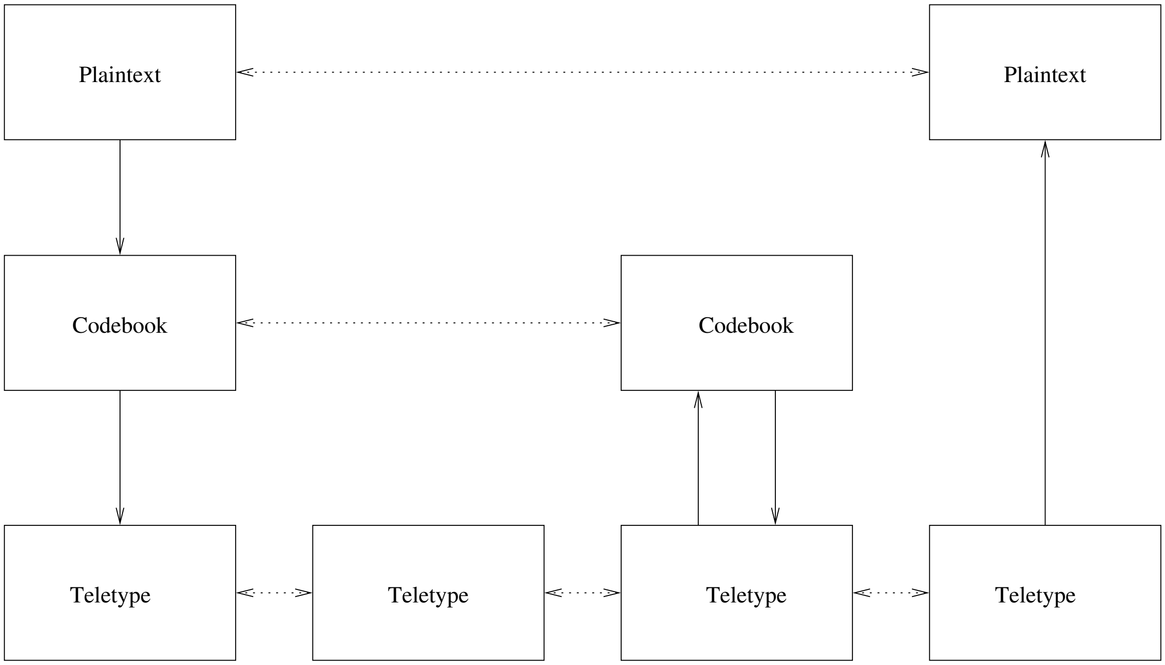


Figure 31: Multi-hop telegram with en route decoding.

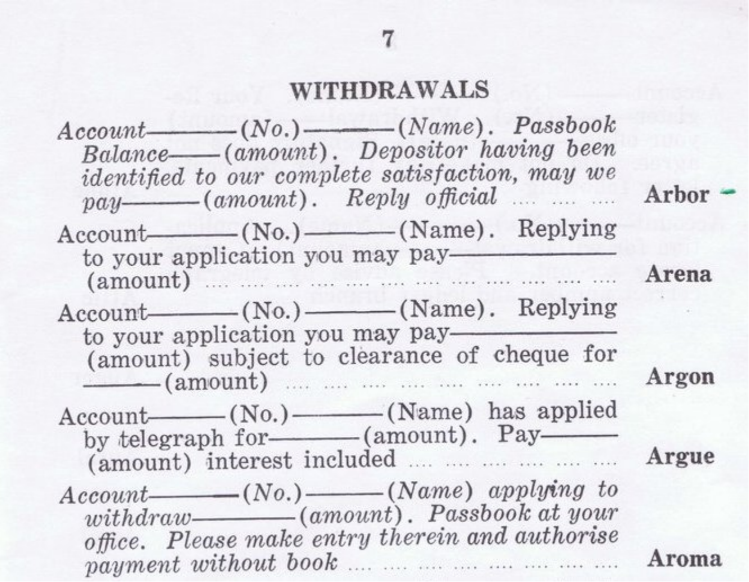
layer. More importantly, lower layer behavior does not change such results. An early, unencoded telegram differed from a traditional sealed letter in that the plaintext was exposed at the transmission layer. Security at that lower layer — for example, selection of a link not easily tapped — did nothing to protect the message from the eyes of the telegraph operator. We see the same thing today with wireless networking: encrypting a single network hop — say, from a laptop to an access point — does nothing to protect the traffic from being intercepted on another hop: there is no end-to-end protection.

It is worth noting that telegrams were also sent via multiple hops. Security protection on one hop, even if perfect, did nothing to protect other hops. Even codebooks were not always end-to-end. Some code companies offered a decoding service. Senders could address their messages to the decoding center; it would produce plaintext and retransmit to the actual recipient (Figure 31). Such a service was useful for transoceanic messages, where the cost of that hop dominated the total cost of the message. Note the peer associations: if encryption is done at the codebook layer, there is no protection against eavesdropping on the link between the decoding station and the recipient. This is analogous to today's virtual private networks (VPN), where traffic from a laptop is encrypted to the organization's firewall but not within the organization.

7 Parting Thoughts

The era of telegraph codes has largely passed. That said, they persisted in some form much longer than is commonly supposed. The Australian Postmaster-General issued a postal banking codebook in 1968 (Figure 32), and the Victorian Railways issued an operational codebook in 1972 (Figure 33). As noted earlier, some types of codebooks are still in use in China for special purposes. The *Manual for Use in Sending Tibetan Telegraphic Wireless Messages* [32] was reprinted as recently as 1985.

It is unlikely that we will glean new technical insights by studying these tomes. What they excelled at has been mathematicized and optimized. That said, the picture painted of the times is still valuable.



7

WITHDRAWALS

Account———(No.)———(Name). Passbook Balance——(amount). Depositor having been identified to our complete satisfaction, may we pay———(amount). Reply official .... **Arbor**

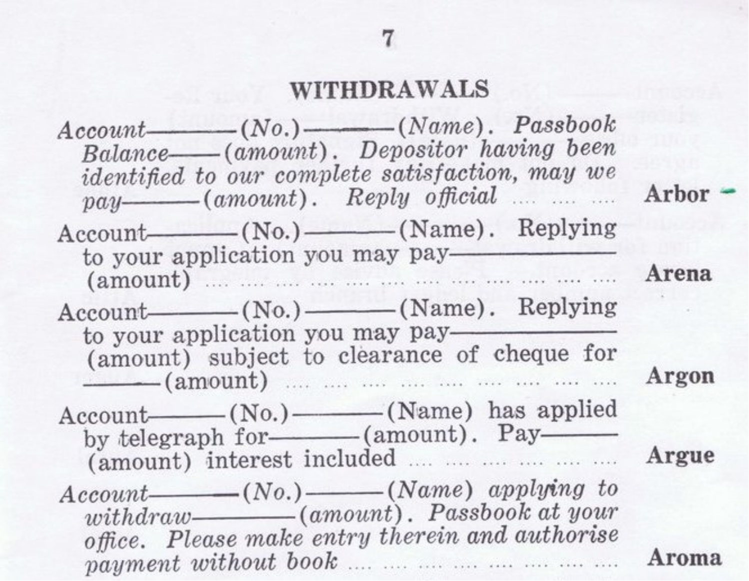
Account———(No.)———(Name). Replying to your application you may pay——————(amount) .... **Arena**

Account———(No.)———(Name). Replying to your application you may pay——————(amount) subject to clearance of cheque for———(amount) .... **Argon**

Account———(No.)———(Name) has applied by telegraph for———(amount). Pay———(amount) interest included .... **Argue**

Account————(No.)———(Name) applying to withdraw————(amount). Passbook at your office. Please make entry therein and authorise payment without book .... **Aroma**

Figure 32: An Australian postal banking code from 1968.



7

WITHDRAWALS

Account———(No.)———(Name). Passbook Balance——(amount). Depositor having been identified to our complete satisfaction, may we pay———(amount). Reply official .... **Arbor**

Account———(No.)———(Name). Replying to your application you may pay——————(amount) .... **Arena**

Account———(No.)———(Name). Replying to your application you may pay——————(amount) subject to clearance of cheque for———(amount) .... **Argon**

Account———(No.)———(Name) has applied by telegraph for———(amount). Pay———(amount) interest included .... **Argue**

Account————(No.)———(Name) applying to withdraw————(amount). Passbook at your office. Please make entry therein and authorise payment without book .... **Aroma**

Figure 33: An Australian postal banking code from 1968.

I close with one final similarity. Today, cryptography is sometimes regulated as a munition. Figure 34, an ad from more than 100 years ago, shows that perhaps the linkage has long been there.

Acknowledgments

> “And yet there lie in his hoards many records that few now can read, even of the lore-masters, for their scripts and tongues have become dark to later men. And Boromir, there lies in Minas Tirith still, unread, I guess, by any save Saruman and myself since the kings failed, a scroll that Isildur made himself.’
>
> —Gandalf, in J.R.R. Tolkien’s *Lord of the Rings*

A paper like this is made immeasurably easier if one has access to a research library, and Columbia University has a superb one. I thank not just the institution but also the librarians and other individuals who have collected and saved obscure works over decades and centuries, against the chance that someone would need them.

I’d also like to thank Jim Reeds. He first showed me commercial codebooks, when I started getting interested in the subject. He also made many useful comments on a draft of this paper, and supplied the error data shown in Figure 7.

Hang Zhao supplied information on modern codebook usage in China and helped with other questions on Chinese codebooks; she and Seung Geol Choi advised me on the Korean Telegraph Code. Ted Lemon provided useful insights on the Tibetan code book. Evelyn Guzman translated the introductory material in [41]. Malek Ben Salem and especially Arezu Moghadam helped with Persian codebooks.

References

[1] Anna Chennault (ed.), *Dictionary of new simplified chinese characters*, Georgetown University. Machine Translation Research Center, Washington, 1962.

[2] Robert M. Coates, *Talk of the town: Rebebureau*, The New Yorker (1934), 16–17. `http://www.newyorker.com/archive/1934/05/26/1934_05_26_016_TNY_CARDS_000237372`

[3] Douglas D. Daye, *A law enforcement sourcebook of Asian crime and cultures: Tactics and mindsets*, CRC Press, Boca Raton, FL, 1997.

[4] William F. Friedman, *The history of the use of codes and code language, the international telegraph regulations pertaining thereto, and the bearing of this history on the Cortina report*, International Radiotelegraph Conference of Washington: 1927 (Washington), United States Government Printing Office, 1928.

[5] William F. Friedman and Charles J. Mendelsohn, *Notes on code words*, The American Mathematical Monthly **39** (1932), no. 7, 394–409. `http://www.jstor.org/stable/2300386`

[6] Great Britain Admiralty, *Nelson’s signals: The evolution of the signal flags*, Printed for His Majesty’s Stationery Office by Eyre and Spottiswoode, London, 1908, Naval Intelligence Division Historical, No. 1.

[7] Richard W. Hamming, *Error detecting and error correcting codes*, Bell System Technical Journal **26** (1950), no. 2, 147–160.

[8] Daniel R. Headrick, *The invisible weapon: Telecommunications and international politics, 1851–1945*, Oxford University Press, New York, 1991.

[9] David Kahn, *The codebreakers*, Macmillan, New York, 1967.

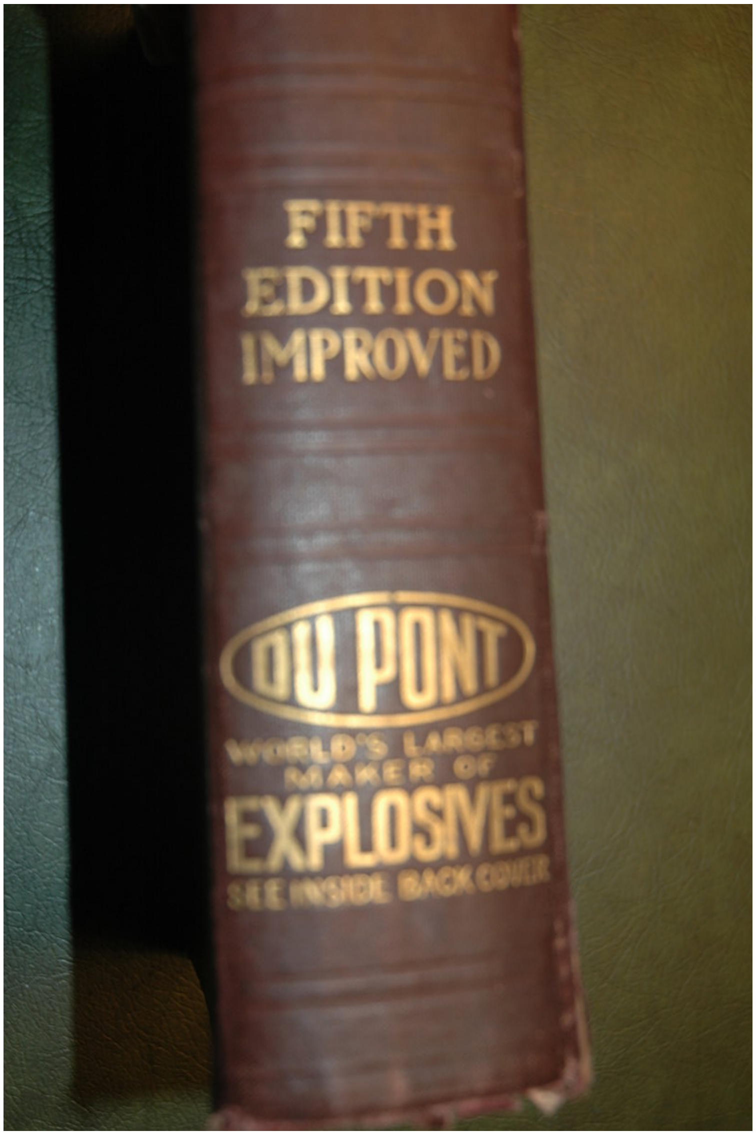


Figure 34: Cryptography and explosives…

[10] ———, *The reader of gentlemen's mail: Herbert O. Yardley and the birth of American codebreaking*, Yale University Press, New Haven, 2004.

[11] Russell Maloney, *Talk of the town: Financial crisis*, The New Yorker (1935), 11–12. `http://www.newyorker.com/archive/1935/11/23/1935_11_23_011_TNY_CARDS_000161388`

[12] G.M. McCune and E.O. Reischauer, *The romanization of the Korean language, based upon its phonetic structure*, Transactions of the Korea Branch of the Royal Asiatic Society **XXXIX** (1939), 1–55. `http://www.nla.gov.au/librariesaustralia/cjk/download/ras_1939.pdf`

[13] W.H. Moody, *Order on use of cipher code in Navy telegraph communications*, 1904, Navy Department General Order 2nd Series No. 153. Location: New York Public Library (microform).

[14] Michael A. Palmer, *Command at sea: Naval command and control since the sixteenth century*, Harvard University Press, Cambridge, Massachusetts, 2005.

[15] Hugh Popham, *A damned cunning fellow: The eventful life of Rear-Admiral Sir Home Popham KCB, KCH, KM, FRS 1762–1820*, Tywardreath, Cornwall, England, 1991.

[16] J. Postel, *Transmission Control Protocol*, RFC 0793, Internet Engineering Task Force, September 1981. `http://www.rfc-editor.org/rfc/rfc793.txt`

[17] Harold Ross, *Talk of the town: Explanation*, The New Yorker (1940), 13. `http://www.newyorker.com/archive/1940/02/17/1940_02_17_013_TNY_CARDS_000179607`

[18] J. H. Saltzer, D. P. Reed, and D. D. Clark, *End-to-end arguments in system design*, ACM Trans. Comput. Syst. **2** (1984), no. 4, 277–288.

[19] Larry J. Sechrest, *Public goods and private solutions in maritime history*, The Quarterly Journal of Austrian Economics **7** (2004), no. 2, 3–27. `http://mises.org/journals/qjae/pdf/qjae7_2_1.pdf`

[20] Tom Standage, *The Victorian Internet: The remarkable story of the telegraph and the nineteenth century's on-line pioneers*, Walker and Co., New York, 1998.

[21] Jonathan Stone and Craig Partridge, *When the CRC and TCP checksum disagree*, SIGCOMM Comput. Commun. Rev. **30** (2000), no. 4, 309–319.

[22] J.R.R. Tolkien, *The lord of the rings*, Allen & Unwin, London, 1954.

[23] Brian Tunstall, *Naval warfare in the age of sail: The evolution of fighting tactics 1650–1815*, Naval Institute Press, Annapolis, MD, 1990, Edited by Nicholas Tracy.

[24] Ralph E. Weber, *United States diplomatic codes and ciphers, 1775–1938*, Precedent Publishing, Chicago, 1979.

[25] ———, *Masked dispatches: Cryptograms and cryptology in American history, 1775–1900*, second ed., Center for Cryptologic History, National Security Agency, 2002. `http://www.nsa.gov/about/_files/cryptologic_heritage/publications/prewii/masked_dispatches.pdf`

[26] Jeanette Williams, *The invisible cryptologists: African-Americans, WWII to 1956*, Series V: The Early Postwar Period 1945-1952, vol. 5, Center for Cryptologic History, National Security Agency, 2005. `http://www.nsa.gov/about/_files/cryptologic_heritage/publications/wwii/invisible_cryptologists.pdf`

[27] Ching yi Dougherty, Sydney M. Lamb, and Samuel E. Martin, *Chinese character indexes*, University of California Press, Berkeley, 1963.

[28] Hang Zhao, 2009, Personal communication.

List of Codebooks

[29] *War Department telegraphic code*, United States Government Printing Office, Washington, 1899–1904. Location: New York Public Library (microform).

[30] *Western Union telegraphic code*, universal ed., International Cable Directory Company, 1900.

[31] *Private telegraphic code of Swift & Company*, Peterson Cipher Code Corporation, 1931.

[32] *Rlung 'phrin gtong deb shes bya kun khyab: Manual for use in sending Tibetan telegraphic wireless messages*, Sambhota Publications, 1985, Reproduced from a rare print of the 1949 Lhasa blocks. Location: East Asian Library, Columbia University.

[33] E.L. Bentley, *Bentley's complete phrase code book*, new ed., London, 1909.

[34] J. G. Bloomer, *Bloomer's commercial cryptograph: A telegraph code and double index—holocryptic cipher*, A. Roman & Co., 1874. Location: Google Books. http://books.google.com/books?id=9OUKAAAAIAAJ

[35] Gilbert E. Brooke, *Code télégraphique AA (the AA cable code)*, second ed., League of Nations Health Organisation, Eastern Bureau, Singapore, 1926. Location: Columbia University.

[36] China Inland Mission, *China Inland Mission private telegraph code*, Methodist Publishing House, Shanghai, 1907. Location: Missionary Research Library, Union Theological Seminary.

[37] ———, *China Inland Mission private telegraph code*, second ed., Methodist Publishing House, Shanghai, 1913. Location: Missionary Research Library, Union Theological Seminary.

[38] ———, *Supplement to the China Inland Mission private telegraph code*, Methodist Publishing House, Shanghai, 1917. Location: Missionary Research Library, Union Theological Seminary.

[39] W. Clausen-Thue, *The A B C universal commercial electric telegraph code*, improved fifth ed., American Code Company, 1915.

[40] Theatrical Code Publishing Co., *The theatrical cipher code; adapted especially to the use of everyone connected in any way with the theatrical business*, Los Angeles, 1905. Location: New York Public Library for the Performing Arts.

[41] Carlos Dobal, *Habla liliś: Un documento secreto*, Biblioteca Nacional, Santo Domingo, 1986.

[42] William Droege (ed.), *The A B C universal commercial electric telegraph code*, sixth ed., Eden Fisher & Co., Limited, London, 1920.

[43] William Droege (ed.), *The A B C universal commercial electric telegraph code*, seventh ed., Eden Fisher & Co., Limited, London, 1936.

[44] Great Britain Admiralty, *Telegraphic signals for the use of His Majesty's fleet*, C. Roworth, London, 1816. Location: EBOOK via Columbia University Library. http://www.columbia.edu/cgi-bin/cul/resolve?clio6396307

[45] Adolphus W. Greely, *Preliminary War Department telegraphic code, supplemental to and to be inserted as an appendix to Western Union telegraphic code*, Government Printing Office, Washington, DC, 1899, War Department Document Number 93. Location: New York Public Library (microform).

[46] Frederic George McCutcheon, *The telegram formula and code combiner*, Marchant SInger and Co., London, 1885. Location: Columbia University.

[47] Bedford McNeill, *McNeill'scode*, 1908 ed., Whitehead, Morris & Co, London, 1908, facsimile reduction.

[48] A.C. Meisenbach, *Acme commodity and phrase book*, Acme Code Company, San Francisco, 1923.

[49] Ernest E. Peterson, *Peterson international code*, second ed., 1923.

[50] Francis O.J. Smith, *The secret corresponding vocabulary, adapted for use to Morse's electro-magnetic telegraph: and also in conducting written correspondence, transmitted by the mails, or otherwise*, Thurston, Ilsley & Co., Portland, ME, 1845. Location: Rare Book & Manuscript Library, Columbia University. `http://books.google.com/books?id=Z45clCxsF7EC`

[51] Richard B. Wynne, *A new code of telegraphic signals for yachts and pleasure boats*, Printed for the author, Edinburgh, 1828. `http://books.google.com/books?id=PC8IGV7fUGMC`

Where a location is noted, it is the physical copy I actually consulted. If no location or URL is given, the book is in a private collection.