CLASSICAL CRYPTOGRAPHY COURSE
BY LANAKI

28 NOVEMBER 1996
Revision 0

LECTURE 20

CODES

SUMMARY

Lecture 20 covers historical use of codes and code systems. We will trace their development and look at some examples of famous code systems. We will develop our subject with the help of several expert references. [FR8], [OAKL], [KAHN], [DEVO], [WEBE], [DEV3], [ELLI], [ACME], [LINC] [SIG2], [DAGA], [TRIT], [MACB], [COLE], [NICH], [MANS], [MAN1], [WEBE]

CODE SYSTEMS

A code system is a highly specialized form of substit bution. The basic principle underling code systems is the replacement of entire words, long phrases, or complete sentences constituting the plain text of a message by arbitrarily selected equivalents having little or no relation to the elements they replace; these equivalents may be other words, groups of letters, groups of figures, or combinations thereof. [FR8]

This replacement process is rarely applied to elements smaller than whole words and when this is done the elements are single letters, groups of letters, or syllables. In a codebook, the words, phrases, and sentences are listed in a systematic manner and accompanied by their code equivalents; correspondents must possess identical copies of the document in order to be able to communicate with each other. An ordinary dictionary may serve the purpose of code communication, so far as single words are concerned, but as a rule a specially prepared document containing the words, phrases, and sentences suited to particular types of correspondence is used. In the U.S. they are called codebooks or codes. Other names come from different locations: repertories, word books, and cipher dictionaries.

TRITHEIM CODE BOOK

One of the earliest code books was developed by the Benedictine Abbot, John Tritheim. He collected many of the ciphers used in the European courts. He was familiar with the occult and proposed a code based on cabalistic words wherein he tried to hide the real meaning under cover of a mysterious language. The courts burned his book "Polygraphia" in great pomp and ceremony. John was lucky that he didn't go with the fire. The first edition was published in Latin in 1518, a French translation in 1541, followed by a German translation.

Part one of Polygraphia consisted of a number of code words for each letter of the alphabet, but arranged in such a manner that if each letter of the message was replaced by a code word, the result was a complete sentence having an innocent meaning. (Something akin to the three column management techno-babble matrix that was popular in the 80's - pick a word from columns A, B and C, put them together and you have a clever sounding and totally bogus phrase like "computer redundancy equivalents'.) Table 20-1 shows the fourteen coded alphabets illustrating the way they were meant to be used:

Table 20-1

TRITHEIM'S CODE ALPHABETS

	1st	2nd	3rd	4th
A	Jesus	Immortal	Producing	Angels
B	God	Omnipotent	Saving	Archangels
C	Saviour	Compassionate	Illuminating	Saints
D	King	Ineffable	Conferring	Spheres
E	Pastor	Universal	Moderating	Heavens
F	Author	Almighty	Expressing	Sea
G	Redempter	Magnificent	Governing	Earth
H	Prince	Puissant	Disposing(of)	World
IJ	Maker	Just	Dominating	Men
K	Conservator	Sempiterneal	Creating	Sun
L	Governor	Celestial	Cognising	Moon
M	Emperor	Divine	Guiding	All
N	Moderator	Excellent	Blessing	Hierarchies
O	Rector	Triumphant	Constituting	Bodies
P	Judge	Clement	Confirming	Spirits
Q	Illustrator	Peaceful	Conducting	Souls
R	Illuminator	Pacific	Sanctifying	Times
S	Consolator	Invisible	Honouring	Humanity
T	Sire	Eternal	Ministrating	Ages
UVW	Dominator	Invincible	Exorcising	Eternity
X	Creator	Benign	Elevating	Firmaments
Y	Psalmist	Pitiable	Sustaining	Stars
Z	Sovereign	Incomprehensible	Vilifying	Air
&	Protector	Excellent	Ordering	Cosmos

	5th	6th	7th	8th
A	Gives	(Tothe) Christians	Eternal	Life
B	Delivers	Requiring (needy)	Perpetual	Joy
C	Attributes	Faithful	Infinite	Joyousness
D	Increases	Attendants	Angelic	Glory
E	Presents	Righteous	Immortal	Consolation
F	Renders	Penitents	Enduring	Felicity
G	Remits	Good	Incomprehensible	Beatitude
H	Renders	Supplicants	Incorruptible	Jubilation
IJ	Envoys	Hopeful	Durable	Tranquility
K	Transmits	Patient	Permanent	Amenity
L	Administers	Afflicted	Ineffable	Recreation
M	Permits	All	Celestial	Clarity
N	Inspires	Tormented	Divine	Union
O	Retributes	Perturbed	Interminable	Peace
P	Orders	Desolated	Perfect	Light
Q	Contributes	Mortals	Sincere	Glorification
R	Frees	Humans	Pure	Benediction
S	Confers	Languishing	Glorious	Security
T	Manifests	Repentant	Supernatural	Favours
UVW	Reveals	Catholics	Indicible	Fruition
X	Maintains	In the World	Peaceful	Happiness
Y	Admits	Sinners	Happy	Light
Z	Agitates	Charitables	Excellent	Exultation
&	Develops	Virtuous	Uplifting	Pleasures

	9th	10th	11th	12th
A	(Together with his Saints)	in Heavens	Majesty	Incomprehen-sible
B	Servants	Ever and Ever	Goodness	God
C	Loved	Without end	Kindliness	Creator
D	Saved	In one Infinity	Sapience	Favour
E	Beatified	Perpetuity	Charity	Jesus
F	Elected	Sempiternity	Power	Transformator
G	Confessors	Enduring	Infinity	Dominator
H	Apostles	Incessantly	Sublimity	Preservator
IJ	Evangelists	Irreversible	Benignity	Immortal
K	Martyrs	Eternally	Commiseration	Supreme
L	Angels	In Glory	Excellence	Mighty
M	Archangels	In the Light	Pity	Omnipotent
N	Dominions	In Paradise	Clemency	Ineffable
O	Proselytes	Always	Mercy	Redemtor
P	Disciples	In divinity	Divinity	Sempiternal
Q	Deified	In Deity	Deity	Governor
R	Ministers	In felicity	Omnipotence	Rector
S	Sanctified	In his reign	Virtue	Sovereign
T	Predestined	in His Kingdom	Love	Invincible
UVW	Preferred	in beatitude	Perfection	Puissant
X	Prophets	in his vision	Force	Merciful
Y	Patriarches	in his magnif-cence	Magnificence	All Powerful
Z	Cherubs	to the Throne	Grandeur	Magnificent
&	Professors	in all Eternity	Favour	Sanctified

	13th	14th
A	Sincerely	Preached
B	Really	Announced
C	Saintly	Published
D	Evangelically	Revealed
E	Devotedly	Denounced
F	Intelligibly	Acclaimed
G	Evidently	Exalted
H	Publicly	Sermoned
IJ	Faithfully	Interpreted
K	Ardently	Reported
L	Constantly	Narrated
M	Sagely	Served
N	Carefully	Praised
O	Virtuously	Recited
P	Catholically	Pronounced
Q	Cordially	Repeated
R	Reverently	Treated
S	Theologically	Speculated
T	Justly	Collated
UVW	Divinely	Spread
X	Learnedly	Cognitized
Y	Entirely	Recognized
Z	Studiously	Contemplated
&	Spiritually	Produced

Amen

Example:

Plain text: 'Do not use bearer.'

```
D O   N O T   U S E    B E  A  R  E  R
1 2   3 4 5   6 7 8    9 10 11 12 13 14
```

Cipher text:

(The) King Triumphant Blessing (the) Bodies Manifests (to the) Catholics Pure Consolation (together with) His Servants (in) Perpetuity The Majesty (of the) Rector Devotedly Treated. Amen.

Look under D in the first alphabet = King, 'O' under the 2nd alphabet = Triumphant, etc.

Note the interesting and rich language in the above 14 alphabets. The unfortunate thing about Tritheim's codes was that coded messages required as many words as there were letters in the plain text, which made for a long cryptogram. Note also that some of the words were duplicated which might have caused some confusion.

FROM LLOYD TO MARCONI

In 1688 Edward Lloyd ran a coffee-house in Tower Street, London. An enterprising man, he found that several brokers used to discuss their business over coffee. To sell more coffee, he decided he must make things easier for them. He instituted a blackboard, and then a weekly bulletin of shipping information. More independent brokers came and consumed his coffee while doing their business. He later moved his coffee house to Lombard street, in the very center of the old city of London frequented by merchants of the highest class. It was not until 1774, with the rapid increase of marine insurance business, a committee was set up and a constitution formed which has remained practically unaltered to the present day. There is no longer a Lloyds' coffee-house, yet the name is preserved, and Lloyds'is known all over the world as the center of the Marine Insurance business.

Lloyds devised a method of signalling between sea and shore, so that advance news of ships and cargoes might be received. A primitive projector was set up and a system of light signals based on the Polybius' system was started. It was this that gave rise later on to the use of codes for commercial purposes; and apart from the Venetian merchants in the eighteenth century, Lloyds signals were the first to come into general use.

In 1794 in Europe, a system of rapid communications known as ' aerial telegraphy', employing semaphores on high towers visible at considerable distances, was instituted. Whole phrases or sentences could be expressed by one group of signals.

In 1825 codes employing figure groups were in common use. In 1845 the Telegraphic Vocabulary Code was used between Liverpool and Holyhead for the semaphore telegraph. In this code there appear words, phrases, long sentences, each represented by groups of one to four digits.

In England the earliest practical trial of electric telegraphy was made in 1837 on the London and North Western Railway, and the first public line, under Wheatstone and Coke Patents, was laid from Paddington to Slough on the Great Western railway in 1843. [DAGA]

In New York, in 1860, Brewell published his Mercantile Cipher for condensing telegrams, in which English dictionary words were employed, and in which we find a fairly complete vocabulary, arranged under captions.

The ABC code, also based on dictionary words, first appeared in 1874. (Refer to Table 20-2.) Up to 1872 the telegraph companies, by international agreement, charged pronounceable code language words as plain text; the higher tariff applied only to cipher or numeral language. These were charged for at a rate of five characters per word; and in 1875, at St. Petersburg, the maximum length was fixed for either plain text or code words at seven syllables. This led to abuse, such as words as Chinesiskslutningsdon - 21 letters, but only 6 syllables - were used by coders. [DAGA]

The rule was changed to apply to European or Latin words but not artificial words. In 1903, code words of ten characters were allowed. They had to be pronounceable to be authorized for transmission at the cost of plain text words.

ABC CODE

Table 20-2

Example of ABC Code Page

Code No.	Half Code Word	Meaning
00000	ABAAA	'ABC' CODE
00001	ABADE	Please use 'ABC' Code 6th edition
00002	ABAEF	Please use 'ABC' Code 6th edition and Code ---- (s)
00003	ABAFG	Please Use 'ABC' Code 6th edition and private Code
00004	ABAGH	Using 'ABC' Code 6th edition
00005	ABAHI	Using 'ABC' Code 6th edition and Code ----------
00006	ABAIJ	Abandon
00007	ABAJK	Abandon altogether
00008	ABAKL	Abandon for the present
00009	ABALM	Abandon or (---)
00010	ABAMN	Abandon the action

In 1904 Whitelaw's Telegraph Ciphers appeared with 400 million pronounceable words. Not really a code book, it was a list of 'artificials' used for private codes. These code words were composed of five letter only, for example FORAB, LUFFA, LOZOJ, etc. as are all words used in commercial codes today. Twenty-thousand words of five letters each were given, and since each was pronounceable, and any two of these words could be joined together to form a chargeable according to telegraph regulations as one word, so 20,000 **2 gave the total of potential words as 400 million.

In 1906 Bentley's code appeared, a compact phrase book based on five-letter groups, applicable to business affairs in general. It cut the cost of international transmissions by half.

MORSE CODE

Samuel Finley Breese Morse was born in 1791 in Charlestown, Mass. His invention of the electrical telegraphy was second only to the famous 'Morse Code'. He based his Morse Code on the frequencies of letters calculated on quantities of type found in the printing office. Since his frequency tables are an enormous help in deciphering every code, lets compare here the original calculation made by Morse with the Normal Frequency and the Telegraph Frequency. (See Table 20-3)

For the letters which were most frequent he used the simplest combination of dots and dashes, which an automatic contrivance of the electric current alternately transmitted and suspended during longer or shorter intervals and reproduced at the other end of the wire on strips of paper. The experienced operator knew the 'fist' of the sender as well as the differences between the dots and the dashes.

Table 20-3

Comparative Table of Order Of Morse's Count with Telegraph Frequencies

	Actual number of letters found by Morse at his printers		Order of Normal Frequency [NICH]
E	1st	12,000	1st
T	2nd	9,000	2nd
A	3rd	8,000	3rd
I	3rd	8,000	6th
N	3rd	8,000	5th
O	3rd	8,000	4th
S	3rd	8,000	8th
H	4th	6,400	9th
R	5th	6,200	7th
D	6th	4,400	11th
L	7th	4,000	10th
U	8th	3,400	13th
C	9th	3,000	12th
M	9th	3,000	16th
F	10th	2,500	15th
W	11th	2,000	17th
Y	11th	2,000	18th
G	12th	1,700	20th
P	12th	1,700	14th
B	13th	1,600	19th
V	14th	1,200	21st
K	15th	800	22th
Q	16th	500	23rd
J	17th	400	25th
X	17th	400	24th
Z	18th	200	26th

Comparative Table of Order of Morse's Count with Telegraph Frequencies

Order	1	2	3	3	3	3	3	4	5	6	7	8	9	9			
Morse:	E,	T,	A,	I,	N,	O,	S,	H,	R,	D,	L,	U,	C,	M			
Telegraph:	E,	O,	A,	N,	I,	R,	S,	T,	D,	H,	L,	U					

Order	10	11	12	13	14	15	16	17	18								
Morse:	F,	W,	Y,	G,	P,	B,	V,	K,	Q,	J,	X,	Z					
Telegraph:	C,	M,	P,	Y,	F,	G,	W,	B,	V,	K,	X,	J,	Q,	Z			

This comparison is remarkable. The normal frequency order corresponds to LANAKI's data presented in Lecture 1. [NICH]

The Morse code was not only used in telegraphy but also in signalling by flags, by flashes of lights, by long and short blasts from a whistle, and for some of us knocks on the wooden cages to fellow prisoners in Viet Nam.

The Army used to allow 10 days for recruit signalmen to learn Morse code. Morse presents a simple method that he invented in Table 20-4. This table presents a short list of words, one for each letter of the alphabet, the long and short syllables indicating dashes and dots.

Table 20-4

Learning Morse Code
(Invented by Morse)

		Morse	Phonetic
A	Ag-ainst	. -	dit dah
B	Bar-ba-ri-an	-...	dah dit dit dit
C	Cont-in-ent-al	-.-.	dah dit dah dit
D	Dah-li-a	-..	dah dit dit
E	(short)	.	dit
F	Fu-ri-ous-ly	..-.	dit dit dah dit
G	Gal-lant-ly	--.	dah dah dit
H	Hu-mi-li-ty	....	dit dit dit dit
I	I-vy	..	dit dit
J	Ju-ris-dic-tion	.---	dit dah dah dah
K	Kan-ga-roo	-.-	dah dit dah
L	Le-gis-la-tor	.-..	dit dah dit dit
M	Moun-tain	--	dah dah
N	Nob-le	-.	dah dit
O	Off-ens-ive	---	dah dah dah
P	Pho-tog-rapher-er	.--.	dit dah dah dit
Q	Queen Kath-er-ine	--.-	dah dah dit dah
R	Re-bec-ca	.-.	dah dit dah
S	Sev-er-al	...	dit dit dit
T	Tea	-	dah
U	Un-i-form	..-	dit dit dah
V	Ve-ry Var-ied	...-	dit dit dit dah
W	Wa-ter-loo	.--	dit dah dah
X	Ex-hi-bi-tion	-..-	dah dit dit dah
Y	Youth-ful and Fair	-.--	dah dit dah dah
Z	(two long, two short)	--..	dah dah dit dit

The famous message SOS = SAVE OUR SHIP = ... --- ... = dit dit dit dah dah dah dit dit dit.

Observe that each of these words contains as many syllables as there are dots and dashes in the corresponding Morse alphabet; but owing to the difficulty of finding suitable words, it was assumed that vowels followed by two or more consonants are long and those by single ones short. In the words Katherine and offensive, for instance, the final syllable must be considered long. Morse put together the following memorization aid:

GALLANTLY and FURIOUSLY he fought AGAINST the foe at WATERLOO.

IVY creeping along the ground suggests HUMILITY.

The JURISDICTION of the NOBLE LEGISLATOR was OFFENSIVE to the BARBARIAN.

A PHOTOGRAPHER saw SEVERAL KANGAROOS on the MOUNTAIN.

U. S. COAST GUARD DISCONTINUES MORSE CODE

Cipher history takes strange turns. It was with some sadness that I read the 31 March 1995 announcement by the USCG that Morse code equipment would shut down after more than 83 years of monitoring telegraph distress calls such as the 1912 Titanic collision with an iceberg. Switchoff occurred at USCG communication centers in Boston, Honolulu, Hawaii, Miami, New Orleans, San Francisco, and Kodiak, Alaska. Even private listening posts will cease service.

Samuel Morse invented the code to carry messages on the telegraph machine he patented in 1840. Morse cipher systems followed soon after that. USCG operators were a breed apart because they could send and receive international language at 20-35 wpm or more. Radio hams know the meaning of "personal touch" as keyed dots and dashes bounce off the atmosphere. WWII vets know that the radioman's "fist" was more identifiable than passwords.

I took my radio training via USCGA SAR. I know the feeling of listening to 11 radios concurrently. Sometimes a May Day could be heard only once and action had to be decisive. Morse has been replaced by automatic equipment to link with Global Maritime Distress and Safety Systems via satellite relayed signals with location fixes.

The ending of the Chesapeake USCG Atlantic COM Center for Morse at 1919 hours EST is the ending of a great era.

I went to my closet today. My USCG web belt still fits. I couldn't bring myself to pull out my old uniform. I just don't look like Arnold Schwartznegger anymore.

COMMERCIAL CODES

Historically, commercial codes were used not so much used for secrecy as for saving money on long telegrams. Authorized, pronounceable words of maximum length of ten letters being used to cover several sentences. The code words used were entirely fictitious, and followed each other in alphabetical order, being made up of five letters each, so that two codewords can be sent by telegraph for the price of a half word. [Note that modern day E-mail on the Net has completely made this a non-issue. In any one day, I may write to classmates in England, Germany, Italy, Japan and Spain and in less than 30 minutes have answers, with attachments, and be charged a flat rate for the service on this end!] Other codes constructed on these principles were Bentley's and Webster's. They allow two words, or even short sentences, to be formed into one telegraph word of ten letters. There are commercial codes today with equivalent translations into every European language, so that English, German, or Italian business men, without knowing each other's languages, can exchange telegrams (or FAXS).

MARCONI CODE

Senator Guglielmo Marconi was devoted to an idea - the sending and receiving of wireless signals through space. His wireless inventions are legendary. Marconi also invented and perfected the Marconi Codes. The complete Marconi code consists of four volumes comprising English, Spanish, Japanese, Russian, Italian, Portuguese, German and Dutch equivalents. The English text is alphabetical, and every other language has a complete index of all the words. The code is divided into two parts - one containing general phrases and the other a numerical system.

Again, the chief aim of standard code was to save cost of cable charges and the cost of time required to code the messages. Upwards of 17,050 combinations could be obtained by the Marconi code. A checking system was used to ensure accuracy.

The code words were composed of five letters each, corresponding to a word or sentence used in trade or business. The codewords could be combined to for a telegraph word of ten letters by the International Telegraph regulations.

There were some differences with codes such as the ABC code. Each code word has a two-letter difference from each other code word. This two-letter difference ensured that no two words would have the same four letters in the same position. A code word like BOPEZ would eliminate codewords like COPEZ, DOPEZ and also such forms as BAPEZ and BEPEZ. (see below)

The Marconi Numerical System was arranged so that a range of figures in combination with some of the most commonly used qualifying phrases, together with an efficiency check, could be transmitted in one complete pronounceable word of ten letters. The first syllable in this section consisted of two consonants, thereby distinguishing it from a phrase section in which none of the code words began win two consonants. As the code words in the numerical section were only two letters long, five words or phrases could be included in one telegraph word of ten letters.

The Marconi arrangement was as follows (Refer to Table 20-5):

1st Syllable provided for a variety of phrases which were employed in combination with the figures or phrases in the following syllables, describing as 'qualifying phrases'; e.g. 'TH' = remit by cable, 'TW' = ship immediately.

2nd Syllable provides for an extensive variety of phrases descriptive of the following weights and measures; i.e. 'OM' =pounds, 'WG' = tons.

3rd Syllable provides for figures from factions to 100.

4th Syllable provides for more figures to be used in conjunction with the third syllable. If unnecessary a blank must be used here, or short phrase to qualify, such as 'ZA' = per month.

5th Syllable provides for a further series of phrases to be used in conjunction with the foregoing; e.g. 'AL' = for immediate shipment. It also supplies a check for the whole coded word.

The checking system is very simple. The check numbers given in brackets on each code syllable are added together for the four syllables used; tens are disregarded, and for the fifth syllable the letters are chosen from the column bearing the same number as the total arrived at from the addition of the first four syllables. Compare the ABC code Table 20-2 with the Marconi code in Table 20-5.

Table 20-5
The First part of the Marconi Code. General Phrases Code words, five letters

Numerical System. Code word of two letters.

No.	Code Word	English	French	Spanish
00000	ABABA	A or an	un, une	un,uno,una
00001	ABAHB	A1 at Lloyds	A1 chez Lloyds	A1 en el registro de Lloyd
00002	ABALC	Abandon(s)	Abandonn(r)(z)	Abandona(r)(u)
00003	ABAND	Abandon all claims	Abandonne toutes reclamations	Abandona todas las reclamaciones
00004	ABAPE	Abandon negotiations	Abandonne les negociations	Abandona las negociaciones
00003	ABARF	Abandon proceedings	Abandonne les demarches	Abandona los procedimientes

1st Syllable.

Check No in Red	Code Syll-able	English	French	Spanish
(0)	BL	Blank or At	Blanc ou A	Blanco o A
(5)	BR	Bid (they)	Ils offrent	ofrecen
(8)	CH	Bid (we)	Nous offrons	ofrencemos
(1)	CL	Bought (we have)	Nous avons achete	Hemos comprado
(6)	CR	Breadth (or thickness)	Largeur (ou epaisseur)	Anchura (o espesura)

2nd Syllable.

(5)	AB	Blank	Blanc	Blanco
(6)	AC	Acre(s)	Acre(s)	Acre(s)
(7)	AD	Ampere(s)	Ampere(s)	Amperio(s)
(8)	AF	Anna(s)	Anna(s)	Anna(s)
(9)	AG	Ante Meredian (A.M.)	Matin, avant midi	Antes de mediodia (A.M.)

3rd Syllable.

(5)	AB	Blank	Blanc	Blanco
(6)	AC	0	0	0
(7)	AD	1/16	1/16	1/16
(8)	BI	1	1	1
(7)	BO	1/14	1/14	1/14

4th Syllable.

(9)	YA	000	000	000
(0)	YB	100	100	100
(1)	YC	200	200	200
(1)	YM	per annum	par an	por ano
(2)	YN	per centimeter	par centmetre	por centmetro

5th Syllable. Control of check.

	0	1	2	3	4	5	6	7	8	9
Blanc	AR	EN	BU	HI	JA	NA	OY	TO	VA	YG
Anout	AC	EP	BY	HO	JE	NE	OZ	TU	YE	YH
Average	AD	ER	CA	HU	JI	NI	PA	TY	VI	YI
C.I.F. (Cost Insurance Freight)	AF	ES	CE	HY	JO	NO	PE	WB	VO	YJ
each	AG	ET	CI	IB	JU	NU	PI	UC	VU	YK

NON-SECRET CODES

Various codes are suited to particular types of correspondence. Many large commercial firms have their own private codes. For example, an early commercial codebook was made by ACME Commercial Code Company in the 1930's. (See Tables 20-2a&b) Most industries have highly specialized technical language (part of the defense of mystique in every industry or profession -Latin for doctors and lawyers, female terms and mathematics for engineers, ISO 9000 terms for quality managers, snake oil terms for computer types, plus a whole bevy of terms for cryptographers, etc). The purposes of many these codebooks are brevity and compression not secrecy. The military and diplomatic applications call for security, and speed of communications, especially for front-line communications.

The PKZIP program, which is used so widely on the net, is a compression 'codebook'. It provides economy of transmission and minimal crypto-security. The power of the program lies in the ability to delineate and hold entire directories and then to create an indexed tree of the coagulated sum of files. PKZIP is an example of a non-secret code. Compression is more valuable than secrecy. The condensing power of a code is dependent on its vocabulary. When we add the goal of secrecy to economy, we then have a secret code. Actually, code transmissions save money because of the lowers number of characters to be transmitted over the channel.

ACME SEVEN DIGIT CODEBOOK

In 1934, the ACME Code Company, with offices in London, New York and San Francisco, developed a codebook for condensation 7 figures into 5 figure groups for international business cables. The transoceanic standard for code language was issued 1 January 1934, and superseded the category 'B' regulation (CDE) five letter code words without vowel restrictions. Category B service was cheaper on short coded messages that the category cable A intercontinental transmissions.

I have codebook number 6015. It is laid out in three tables on each page. Table 1 is for 1st and 2nd Figures, First and Second Letters; Table 2 is for 3rd, 4th and 5th Figures, Third and Fourth letters; Table 3 is for 6th and 7th Figures, fifth letters. The conversion (condensation) of 7 Figures into a five letter code word and visa versa is accomplished on one page, in a single operation. Numbers 0000000 to 9999999 are included in the codebook. The condenser is used for :

14 Figure codes (2 Five Letter code words)
21 Figure codes (3 Five Letter code words)
28 Figure codes (4 Five Letter code words)
etc.

It can be used in conjunction with any numbered code, catalogue, parts list, steamer list, etc.

Encoding

To encode the figures 4732651 into a five letter code word, we divide the 7 figures into three groups:

47 - 326 - 51

The 3rd, 4th, and 5th figures determine the page from which we apply the condensation codes. 326 is found on page 3 of the codebook. (Table 20-6a&b reproduces part of the pages.)

Alongside the figures 326 are the two letter groups YM, YN, YO, YP, which gives us the 3rd and 4th letters of the codeword we will form.

To determine the group to use, we look at the 6th and 7th figures table and find the fifth letter. The figures 51 are in the same column as YP and the letter alongside of 51 is M. Thus we have the 3rd, 4th, and 5th letters of our codeword YPM.

To get the 1st and 2nd letters of the codeword, we refer to the table covering the 1st and 2nd figures, on the same page and is found that for 47 are FR. The entire codeword is then FRYPM.

Table 20-6a
page 3

3rd, 4th, and 5th Figures
Third and Fourth Letters

326	YM	YN	YO	YP
327	YQ	YR	YS	YT
328	YU	YV	YW	YX
329	YY	YZ	ZA	ZB
330	ZC	ZD	ZE	ZF

.	.	.	.		
			39	J	
6th and			43	K	
7th			47	L	
Figures			51	M	
Fifth			55	N	
Letter			59	O	
.	.	.	.	.	.

..	..	
45	FP	
46	FQ	
47	FR	1st and 2nd
48	FS	Figures
49	FT	First and
50	FU	Second Letters
..	..	

The codeword then is FRYPM.

Decoding

To decode the codeword STROW we break it down:

ST - RO - W

The first and second letters determine the page from which you will decode your full seven figures. In this instance the First and Second letters ST, they will be found on page 6, alongside of which we find the 1st and 2nd figures "87".

You then look on the same page for RO, in the table covering the third and fourth letters. It is found that RO means "782," thus giving you the 3rd, 4th and 5th figures.

W being the final letter, we refer to the table for 6th and 7th figures. In the same column as RO appear and on the same line that the letter W is found, we find the 6th and 7th figures 84. Our final product is 87-782-84. STROW = 8778284.

Table 20-6b
page 6

3rd, 4th, and 5th Figures
Third and Fourth Letters

	782	RO	RP	RQ	RR
	783	RS	RT	RU	RV
	784	RW	RX	RY	RZ
	785	SA	SB	SC	SD
	786	SE	SF	SG	SH
		.	.	.	.
6th and		84	85	86	87	W
7th		88	89	90	91	X
Figures		92	93	94	95	Y
Fifth		96	97	98	99	Z
Letter		..	..	..	..	.

..	..
85	SR
86	SS
87	ST
88	SU
89	SV
90	SW
..	..

1st and 2nd Figures
First and Second Letters

ACME also produced a Commodity and Phrase Code Supplement which was just as much fun? [ACME]

BREVITY CODES

In military cryptography, the greatest degree of condensation is afforded by 'prearranged-message codes,' or 'brevity codes.' A prearranged-message code is a tactical code adapted to the use of units requiring special or technical vocabularies; it is comprised almost exclusively of groups representing complete or nearly complete messages and is intended for shortening messages and concealing their content. The police '10' codes fall into this category. A brevity code has as its sole purpose the shortening of messages. A field code is a small tactical code which contains a large number of code groups representing words and a few common short phrases, from which sentences can be composed; a syllabary, which is a list of code groups representing individual letters, combinations of letters, or syllables, is used for spelling out proper names and . numerical tables, or list of code groups representing numbers, dates, and jargon. The Army Special Forces Codes fall into this category. A jargon code is a very short code in which bona fide dictionary words, baptismal names, rivers, lakes, etc are used as code groups. Lincoln's war time codes fall into this category. [LINC] A voice code or recognition code is used for transmission by small radio-telephone sets used in combat. Other names are combat code or operations code. [TEC] The Navy has a special brand of codes used for protection of marine traffic. An example of this code system is the International Code of Signals (1969 edition, revised 1981 INTERCO) [SIG2]

INTERNATIONAL CODE OF SIGNALS FOR VISUAL, SOUND AND RADIO COMMUNICATIONS (INTERCO)

The Defense Mapping Agency, Hydrographic/Topographic Center issued in 1969 and again in 1981, their Publication No. 102, "International Code of Signals For Visual, Sound, and Radio Communications," United States Edition. This code was adopted by the Fourth Assembly of the intergovernmental Maritime Consultative Organization in 1965. The document was prepared in nine languages: English, French, Italian, German, Japanese, Spanish, Norwegian, Russian and Greek.

This is very good example of the brevity and non-secret codes that had wide distribution for ocean going vessels. Modern day vessels use uplinks to satellites in geo-synchronous orbits to navigate and communicate.

The INTERCO was designed to communicate for situations relating to the safety of navigation and persons, especially when language difficulties arise. It is suitable for transmission by all means of communication including radiotelephony and radiotelegraphy. The INTERCO embodies the principle that each signal has a complete and distinct meaning.

The INTERCO is broken into four parts: 1) Signal Instructions, 2) General Signal Code, 3) Medical Signal Code, and Distress and Lifesaving Signals and Radio Procedures. The appendix includes a National Identity Signals for Ships and Aircraft, plus US/USSR Supplementary Signals for Naval Vessels.

General Signal Code includes sections on: Distress, Emergency, Casualties, Damages, Aids to Navigation, Hydrography, Maneuvers, Cargo, ballast, Meteorology, Communications and Sanitary Regulations. [SIG2] See Table 20-7 for sample entries. In Table 20-7, capitalized headings represent major topics, predominantly lower case headings represent subtopics. You can see from the small sample in Table 20-7, that the INTERCO deals with serious situations. I was assigned to a U.S. Coast Guard Radio Room and I can tell you that listening to 11 radios at the same time can be very intense. A MAYDAY maybe heard only once and rarely in calm voice. Sending the cutter is serious business. The USCG does their job exceptionally well.

Table 20-7
Sample Entries from INTERCO Codebook

Distress - Emergency

Code	Meaning
	ABANDON
AD	I am abandoning my vessel which has suffered a nuclear accident and is a possible source of radiation danger.
	Accident
SB	I am proceeding to the position of the accident.
GC 2	I have searched area of accident but have found no trace of derelict or survivors.
	Doctor
AM	Have you a doctor on board?
AP	I have ... (number) casualties.
	ASSISTANCE
	Required
CB	I require immediate assistance.
CB 1	I require immediate assistance; I have a dangerous list.
CB 6	I require immediate assistance; I am on fire.
	Given-Not Given
CN 1	You should give immediate assistance to pick up survivors
CO 1	I cannot give the assistance required (or vessel/aircraft indicated)
	DISABLED-DRIFTING-SINKING
DS	I have sighted disabled aircraft in lat ... long ... at time indicated.
DX	I am sinking.

SEARCH AND RESCUE

Proceeding To Assistance

FE I am proceeding to the position of the accident at full speed. Expect to arrive at time indicated.

Position of Distress or Accident

FF I have intercepted SOS/MAYDAY from vessel (name or identity signal or aircraft) in pos lat ... long ... at time indicated.

Results of Search

GJ 1 Wreckage is reported in lat .. long ... No survivors appear to be in vicinity.

ICEBREAKER

WC 1 Icebreaker is being sent to your assistance.

SEA

WY The state of the sea is ... (Complements 0-9 corresponding to following table):

	Height In Meters	Height In Feet
0 Calm (glassy)	0	0
1 Calm (rippled)	0 - 0.1	0 - 1/3
2 Smooth (wavelets)	0.1 - 0.5	1/3 - 1 2/3
3 Slight	0.5 - 1.25	1 2/3 - 4
4 Moderate	1.25 - 2.5	4 - 8
5 Rough	2.5 - 4	8 - 13
6 Very Rough	4 - 6	13 - 20
7 High	6 - 9	20 - 30
8 Very High	9 - 14	30 - 45
9 Phenomenal	over 14	over 45

MEDICAL

Diseases of Respiratory System

MIF Patient is coughing up blood.
MIM Patient has blueness of face.

Special Treatment

MRW Give frequent gargles one teaspoon of salt in a tumblerful of water.

RECEPTION OF SAFETY MESSAGES

MAYDAY (Distress) Indicates that the ship, aircraft, or other vehicle is threatened by grave and imminent danger and requests immediate assistance.

PAN (Urgency) Indicates the calling station has a very urgent message to transmit concerning the safety of a ship, aircraft or other vehicle, or the safety of a person.

SECURITE (Safety) Indicates that the station is about to transmit a message concerning the safety of navigation or giving important meteorological warnings.

To indicate DISTRESS:

1. If possible transmit ALARM SIGNAL (i.e. two tone signal) for 30 seconds to one minute, but do not delay the message if there is insufficient time in which to transmit the Alarm Signal.

2. Send the following DISTRESS CALL:

MAYDAY MAYDAY MAYDAY. This is ...(name or call sign of ship spoken three times).

3. Then send the DISTRESS MESSAGE composed of:

MAYDAY followed by the name or call sign of the ship;
Position of ship;
Nature of distress;
And if necessary, transmit nature of the aid
required and any other information which will help
the rescue.

USE PLAIN LANGUAGE WHENEVER POSSIBLE or send the word INTERCO to indicate that the message will be in the International Code of Signals.

example:

MAYDAY MAYDAY MAYDAY ... (name of ship spoken three times, or call sign of ship spelled using Phonetic Alphabet in Table 20-8); MAYDAY ... (name or call sign of ship) Position 54 25 North 016 33 West I am on Fire and require immediate assistance.

Table 20-8
Phonetic Alphabet used with INTERCO

Letter/ Number	Word	Pronounced
A	Alfa	AL FAH
B	Bravo	BRAH VOH
C	Charlie	CHAR LEE or SHAR LEE
D	Delta	DELL TAH
E	Echo	ECK OH
F	Foxtrot	FOKS TROT
G	Golf	GOLF
H	Hotel	HOH TELL
I	India	IN DEE AH
J	Juliett	JEW LEE ETT
K	Kilo	KEY LOH
L	Lima	LEE MAH
M	Mike	MIKE
N	November	NO VEM BER
O	Oscar	OSS CAR
P	Papa	PAH PAH
Q	Quebec	KEH BECK
R	Romeo	ROW ME OH
S	Sierra	SEE AIR RAH
T	Tango	TANG GO
U	Uniform	YOU NEE FORM or OO NEE FORM
V	Victor	VIK TAH
W	Whiskey	WISS KEY
X	Xray	ECKS RAY
Y	Yankee	YANG KEE
Z	Zulu	ZOO LOO
0	NADAZERO	NAH-DAH-ZAY-ROH
1	UNAONE	OO-NAH-WUN
2	BISSOTWO	BEES-SO-TOO
3	TERRATHREE	TAY-REE-TREE
4	KARTEFOUR	KAR-TAY-FOWER
5	PANTAFIVE	PAN-TAH-FIVE
6	SOXISIX	SOK-SEE-SIX
7	SETTESEVEN	SAY-TAH-SEVEN
8	OKTOEIGHT	OH-TAY-AIT
9	NOVENINE	NO-VAY-NINER
.	DECIMAL	DAY-SEE-MAL

BASICS OF CODE CONSTRUCTION

The encoding and reverse procedure of decoding is accomplished by replacing various words, phrases, sentences, and numbers by their code equivalents. The code text is built up from code units each representing the longest possible plaintext unit the code book affords. Encoding the phrase "enemy force estimated at one battalion," and the codebook has phrases "enemy force," and "estimated at," as well as the individual words, we would write down the phrase equivalents.

The elements of which code groups are composed may be one or more of the following:

1. Bona fida words - real words from Dutch, English, French, German, Italian, Latin, Portuguese and Spanish.

2. Artificial words - groups of letters without meaning with vowels and consonants arranged to appear like real words.

3. Random groups of letters.

4. Groups of Arabic figures.

5. Intermix groups, ie. call signs for stations K2KAA, or W5AZZ.

6. All the above.

PARALLEL SETS

A code may contain two or more parallel sets of code groups of different types. In many commercial codes and some military codes, there is one series of code groups of the bona fide type or artificial word type and another series of the figure-group type, both applying to the same series of words phrases, and sentences of the code. In parts of the world where English letters are used for writing, letters possess greater advantages in accuracy of reading than figures - especially for telegraph or radio transmissions. For communications to China and Russia or obscure ports, Arabic figures are well accepted and code groups composed of figures are used. The main reason for this is assurance of the correct transmission and reception of messages in all parts of the world. Another reason is that certain methods of enciphering code messages for the sake of greater secrecy, figure groups often form the basis for encipherment more readily than do letter groups.

The greatest advantage possessed by letter groups over figure groups lies in the availability of a far greater number of permutations, or interchanges, of letter groups, because there are 26 letters which may be permuted to form letter groups compared to 10 digits for figure groups (assumes base 10 historical use). If code groups of five letters are used, then there are 26 ** 5 or 11,881,376 groups of five letters versus 10 ** 5, or 100,000 groups of five figures. Letter code groups are usually constructed to reduce error in transmission.

The length of code groups used, whether the groups consist of two, three, four, or five elements, dependsupon the size of the code. This applies almost exclusively to field military or naval codes, where transmission is through a governmental agency; in commercial messages or governmental communications transmitted over privately operated lines, five-letter or five letter groups are the standard. [FR8]

Code groups of modern codes are constructed by the use of tables which permit more-or less automatic and systematic construction in the form desired. These are called permutation tables. Because they may be used to correct most errors made in transmission or writing, such tables are usually included in the code book and are called mutilation tables, garble tables, error detector charts, etc.

TWO-LETTER DIFFERENTIAL

The average telegraph or radio operator did not work without error. One letter different code groups like ABABA and ABABE were easy to mistake and the message could be made unintelligible by only a few transmission errors. If however, every code group in the code book is distinguished from all other code groups in the same code by a difference of at least two letters, then there would have to be two errors in a single group and these two errors would have to produce a code group actually present in the code before a wrong meaning would be conveyed. The principle of making code groups differ by a minimum of two letters is called the two- letter differential. The two-letter differential reduces the possibilities for constructing letter code- groups from 26 ** 5 to 26 ** 4 (456,976) but considering the advantages, the sacrifice was worthwhile. Permutation tables for construction of figure-code groups are similar in nature

and purpose to tables for construction of letter-coded groups. Because of a more limited number of characters available for permutations, the maximum number of 2-figure difference groups possible in a 5-figure code is 10 ** 4, or 10,000. (This does not account for ASCII code derivations.)

TYPES

In their construction or arrangement, codes are generally of two types:

(1) One-part, or alphabetical codes. The plaintext groups are arranged in alphabetical order accompanied by their code groups in alphabetical or numerical order. Such a code serves for decoding as well as encoding.

(2) Two-part or randomized codes. The plaintext groups are arranged in alphabetical order accompanied by their code groups in a non-systematic order. The code groups are assigned to the plaintext groups at random by drawing the code groups out of a box in which they have been thoroughly mixed. Such a list serves for encoding. For decoding, another list must be provided in which the code groups are arranged in alphabetical or numerical order and are accompanied by their meanings as given in the encoding section. Another name for the two-part code is cross- reference codes. Here are extracts from typical one-part and two-part codes. (Tables 20-9 and 20-10.)

Table 20-9
One-part code

ABABD	A
ABACF	Abaft
ABAHK	Abandon
ABAJL	.....it
ABALN	Abandoned
ABAMP	.....by
ABAWZ	Abandoning
ABBAD	Abandonment
.....	.....
.....	.....
ZYZYZ	Zero

Table 20-10
Two-part code

Encoding Section		Decoding Section	
GAJVY	A	ABABD	Obstructed
TOGTY	Abaft	ABACF	Term
FEHIL	Abandon	ABAHK	Zero
BAYLT	.....it	ABAJL	If it has not
ZYZYZ	Abandoned	ABALN	To be sent by
NYSYZ	.....by	ABAMP	Acceding
IFWUZ	Abandoning	ABAWZ	Building
RUMGO	Abandonment	ABBAD	Do not attempt
.....		.....	
.....		.....	
ABAHK	Zero	ZYZYZ	Abandoned

Between the two extremes are codes which have features of both; that is complete sections may be arranged in random sequence, but within each section the contents are arranged in some logical order.
When a strict alphabetic arrangement is used in the sequence of the phrases, the code is said to be a strictly alphabetical code. When the phrases are listed under separate headings based upon the principal word or idea in the whole expression, the code is called a caption code. (Tables 20-11 and 20-12)

Table 20-11
Caption code

```
Assistance
     Give assistance
     Require assistance
     No assistance
     Assistance has been sent
     Assistance for
     Assistance from
     Assistance to
Assistant
Assisted
```

Table 20-12
Strictly-alphabetical

```
Assistance
     Assistance for
     Assistance from
     Assistance has been sent
     Assistance to
Assistant
Assisted
     ........................

Give
   Give assistance
     ........................
No
   No assistance required
     ........................
Require
   Require assistance
```

More precise and economical coding is possible with a caption code than with an alphabetical code. With a caption code it is easier to assemble an extended variety of expressions and shades of meaning under specific headings than with alphabetical code. On the other hand, the use of a caption code involves more time and labor in encoding.

Two-part codes are used by many governments for their secret diplomatic, military and naval communications because of the advantages they offer over one part codes. Some disadvantages include twice as large in context, printing and distribution costs, compilation is four times greater because of the requirement of accurate cross references. The advantages of two-part codes are greater security and greater accuracy.

In some commercial code messages there is sometimes encountered the practice of mixing plaintext and code text. In governmental and naval communications such intermixtures are rare and present an abysmal ignorance of the fundamental rules of cryptographic security. Because the plaintext words give definite clues to the meaning of the adjacent code groups, even though the former convey no meaning in themselves (such words as and, but, by, comma, for, in, period, stop, that, the, etc) constitutes a fatal danger to the message security.

ENCIPHERED CODE SYSTEMS

Sometimes the code groups of a code message undergo a further process of encipherment; the resulting cryptogram constitutes an enciphered code message. Both transposition and substitution may be used to encipher the code. Enciphered code is used under the following circumstances:

(1) When the code has a wide distribution and may fall into enemy hands,

(2) to improve the security of commercial codes and nonsecret codes, and

(3) when increased security is necessary for highly classified communications.

Transposition methods are generally used within code groups, such as rearranging or shifting about the letters or figures composing them. A common method is keyed columnar transposition with special matrices with nulls. All the substitution methods previously studied may be used for "super-encipherment" of the code. The most effective methods of enciphering code are arithmetical methods.

If the code groups are numerical, the addition (usually mod 10) of an arbitrarily selected number (called the additive) to each code group message constitutes a simple form of encipherment. The additive may be fixed.

Additive methods may actually be weak cryptographically if the basic code book and code groups embody limitations in construction. Instead of adding a fixed number in encipherment, the latter is subtracted, in which case , in decipherment, the fixed number must be added to the enciphered code groups as received. Such a group (called subtractive or subtractor) in decipherment the group becomes an additive. A third method used commonly is the minuend method. It involves the subtraction of the plain code group from the key to yield the enciphered code group in encipherment, and the subtraction of the enciphered code group in from the key in decipherment. Addition and subtraction of a fixed numerical group may be alternated within the same message such as +200, +100 +400 as a cycle or +200, - 100, +400, -200 etc. Instead of a fixed additive, it is possible to employ a repeating large key.

When special tables are employed as the source of the additives or subtractors for enciphered code, a much more secure system is provide. These tables are called a key book or an additive book or a subtractor book. by applying identifying symbols called indicators to the pages, as well as to the rows and columns on each page of the key book, it is possible to provide for secure encipherment of a large volume of traffic. All corespondents must have the same key books. In employing the key book, the indicators tell the recipient of the message what key groups were used and where to begin the decipherment of the enciphered code.

In actual practice, indicators are often disguised or encrypted by a special key or set of keys; this procedure may add considerably to the security of the system.

Table 20-13 shows a page from a typical key book. It contains two sets of 100 4-digit key groups, disposed in numbered blocks each containing 10 rows and 10 columns of groups. To designate a group as the initial one to be employed in encipherment or decipherment, we give the block number, the row and column numbers of the group. For example, 0116 is the indicator for the group 8790. It is usual to take the successive groups in the normal order of reading. Some keys books consist of 50 + pages containing 200 + groups making 10,000 in all. The digits in each block are random numbers. [FR8] If the key book is used once and only once, security of the system approaches the one-time pad. The messages are one time system secure even if the enemy has basic code book. Friedman discusses indicators in much more detail in [FR8].

Table 20-13
Indicators and Key Blocks

Block 00

	1	2	3	4	5	6	7	8	9	0
1	0378	9197	3260	3607	2699	9053	9733	1844	6622	4213
2	7185	0135	6091	2387	4957	3113	7284	0750	3501	1945
3	5037	3365	1294	8261	2149	0718	3678	2510	7238	5268
4	8004	5199	3859	1293	5311	3550	9915	0512	1518	3776
5	9282	6893	4229	9736	0927	1418	1930	9864	0090	8974
6	7259	9399	0769	3144	9801	1378	4732	5134	1435	5282
7	2878	9963	7943	4519	3404	9810	1090	4467	7069	5348
8	1620	5879	0218	1064	9560	5732	6661	0883	1883	2619
9	3868	1905	2500	6654	0824	3710	3875	6332	1503	7259
0	4319	3298	7819	8721	1549	6630	6301	5701	3586	1907

Block 01

	1	2	3	4	5	6	7	8	9	0
1	9328	1135	3871	1549	0839	8790	1771	8251	3274	1173
2	2297	9550	5033	0102	6817	5579	0847	4038	1200	2949
3	3640	3984	3299	1181	3811	8844	2500	4557	4133	0487
4	1256	9614	5520	8372	1941	2417	1098	4039	3943	8282
5	1751	4254	8479	8647	2684	5511	8680	4660	2315	4857
6	4587	5968	2568	1254	0258	1254	3568	2548	4521	8795
7	1258	6241	0125	2458	4587	5632	2589	1548	1235	1458
8	1254	2548	0004	4561	2565	2437	7849	1245	3265	4879
9	4582	1546	2589	2145	7854	7895	4589	6369	3698	1254
0	1255	1544	7850	2569	9989	8754	2548	1220	0387	0589

DICTIONARY CODES

Dictionary codes are highly specialized forms of substitution systems. Code books (modified dictionaries) used by the Department of State and military represent a greater condensation of words than commercial systems - a single code group may represent a long phrase. The average condensation of a diplomatic code is 1:5 while a commercial code is only 1:3. [DAGA] By way of comparison, modern PKZIP compression is 1:3 - 1:4 on normal text. I recently experimented with PKZIP on the TEA program library for eight words and up and found an average compression of 1:2.5. These groups are all pronounceable artificial words. For example, we might have ABACA in commercial code, EXA in diplomatic code and occasionally syllables as BA in Marconi code.

It is difficult to safeguard against the loss of codebooks which have to be printed in fair numbers. Macbeth reports on an interesting story about Ottoman Field-Marshal Osman Pasha during the Russo-Turkish war in 1877. Pasha entrusted one of his generals, Selim Pasha with a confidential mission. Selim was the officer in charge of ciphers and codes and always kept the code book on his person. Selim departed so promptly on his mission that he forgot to leave the volume with his chief. And the latter, during the whole time of the Adjutant's absence, saw a pile of ciphered messages from Constantinople accumulate on the table without being able to read or reply to them. [DAGA]

Codes used in conjuction with ciphers (superencipherment) can be very difficult to break; but the work and time involved in making this combination can be significant (if done by hand in the field.) Computers reduce the legwork significantly.

The typical dictionary code protocol is as follows:

1) Agree with the recipient on the exact edition of the diction to be used, i.e. Concise Oxford Dictionary, current edition, by Fowler and Le Mesurier.

2) Use the number of the page, and the number of the word down the page to encipher:

Given Plain: " Reunion Berlin Tomorrow"

Code:

1006 (page no.), 12(word no) = Reunion
0104 (pages with fewer than four numbers would have a 0
added in front to keep to the uniformity), 17 (word no.) = Berlin
1291 - 08 (on the same principles) = To-morrow

Ciphertext:

100612 010417 129108

These figures, if greater secrecy is required, could again be enciphered and thus converted into letters by means of an agreed upon cipher.

3) Prepare for superencipherment by dividing the figures into pairs and then convert them into letters by means of a table such as Table 20-14.

Table 20-14

Digraphic Equivalents for Superencipherment

	1	3	5	2	4	9	7	8	6	0
9	AN	DA	HN	JT	MB	KC	GF	ES	BZ	ZA
2	CK	AO	DB	HO	JS	GE	ER	BY	FR	YB
7	IR	CJ	AP	DC	GD	EQ	BT	FQ	LH	VA
4	MC	IY	CI	AR	DD	BS	FP	LI	NL	VB
8	MA	KB	GC	CG	AS	DF	HP	JU	OB	VC
1	KA	GB	EP	BR	CE	AT	DG	HQ	JQ	TZ
5	GA	EO	BP	FO	IX	CC	AX	DH	HR	TY
3	EN	BO	FN	LJ	NK	IZ	CB	AY	DJ	SB
6	BN	FM	LK	NJ	OA	OC	IV	CB	AZ	QA
0	XY	YA	BY	YB	XC	XE	YD	YE	YX	QC

Nulls: WA WE W, to end message in groups of fove letters.

The numbers enciphered into letters:

TZYXBR XYXCDG BRANYE

and the cryptogram for transmission:

TZYXB RXYXC DGBRA NYEWA

The suggested cipher can easily be arranged to make pronounceable words suitable for telegraph or radiotelegraph transmission.

Certain dictionaries have been issued which give two columns on each page with words directly opposite to each other. Then it is possible to give the word opposite the one we really mean, or a word which is 5 or 3 or 10 places either above or below the one we want to encode. Codes of this kind can be solved readily.

CRYPTANALYSIS OF A SIMPLE DICTIONARY CODE

An Australian criminologist named Mansfield presented some interesting principles for solving dictionary codes. He calculated dictionary progressive lists, giving numbers of words beginning with any two letters in dictionaries of 10,000 - 100,000 words. [DAGA]

Given:

55381 42872 35284 44381 45174 56037 55381 46882
23171 44234 55366 55381 00723 12050 61571 36173
55381 56442

We rearrange the list from lowest numbers to highest.

00723 42872 55381 (5 times)
12050 44234 56037
23171 45174 56442
35284 46882 61571
36173 55366

Words beginning with XYZ are seldom used, so we can take it that the highest number indicates a word beginning with a W or a T. [Mansfield made big assumptions about nulls and standardization of the dictionary. Lectures 2 and 3 showed how we can rip this assumption to shreds.] But the list of bigram frequencies (from Lecture 1) gives us the commonest initial group as TH or THE, and if we fix any repetition of such nature, then we may have the T in that dictionary. Naturally, we start with 55381 occurring five times and assume it is THE.

The highest number after that is 61571, so that it could indicate a word beginning with a W. This gives us a clue to the probable number of words in the dictionary used for the code. It cannot be over 65,000 words as XYZ words are very few, seldom more than 3,000. [This part of Manfield's analysis is an extraordinary jump of faith -what is more extraordinary is that it will work more than 60% of the time on simpler dictionary codes.]

According to Mansfield's Progressive Dictionary Lists, we attempt to fix the probable first two letters of each word in the code. For instance the 2nd group 12050 will be between 11646 (terminating words beginning with DA) and 12850 (terminating words beginning with DE), so that it is probable to be a word beginning with DE. [DAGA] [MANS], [MAN1]

Using Mansfield's lists we obtain:

THE RE--- OF THE RO--- TO- THE SE- -HA - RE- TH- THE
RE- DE- - WA- OV- THE TO-

We locate in the dictionary the word THE (55381) and count back twenty words for 55366 (th). This gives us an area covering words THANE, THANK, THAT, THATCH. We try the most likely THAT. We note the two words starting with letters TO- 56037 and 56442. Words beginning with TO start at 56037 and stop at 56466, so that it is reasonable guess to assume the first is TO and the second (56442), we count twenty words back to find the word TOWN.

The R group is: -RE- (42872) and RE- (44234) and RO- (45174). RE stands 300 words from the end of the RA's which stop at 42573, according to Mansfield's tables. This gives us the following words to select from: RECLINE, RECOMMEND, RECOMPOSE, RECONNAISSANCE, RECOUP, and RECOVER. We choose RECONNAISSANCE. The next look at our cipher is:

THE RECONNAISSANCE OF- THE ROUTE TO THE SE- HAS-
REVEALED THAT THE AE- DE- WA- OV- THE TOWN.

We apply the same process to the AE- 00723 and get airplane, while the DE- 12050 occurring one-quarter of the way from the end of the DA to the end of the DE brings us to DEF, limited by DEFACE and DEFY, where only DEFEAT, DEFENSE, DEFEND, and DEFENSIVE are probable. We select airplane defensive us near the mark.

SE- should be sea 46882 and OVER for OV- 36173. The of- is in fact OF, and the HA- is has, and the WA- is was. The complete message reads:

THE RECONNAISSANCE OF THE ROUTE TO THE SEA HAS
REVEALED THAT THE AIRPLANE DEFENSIVE WAS OVER THE
TOWN.

[MANS] tells us that the real message was off by two words. Instead of AIRPLANE DEFENSIVE, it was AIR DEFENSES, but the meaning was essentially the same.

What Mansfield did show us in 1936 was that the laws of probability work with dictionary codes. The search in the area of possible words will give us the root of the plain text so that we may deduce the whole meaning of the code.

DIPLOMATIC CODES

One of the best references on historical codes (1775-1938) in the United States was written by Professor Ralph Weber. [WEBE] He describes one interesting code used in 1867 by the State Department known as WE029. (Refer to Table 20-15) It used a simple substitution masking procedure, eliminated the use of the letter W because it was not used in European or Latin nations, focused on 24 letters of the alphabet and assigned them to the 24 most common parts of speech such as articles and other words (s= plural; a = THE; e = AND, etc.) Other ordinary words were assigned to the approximately 600 combinations of 2 of the letters. Three letters were used for the remainder of the vocabulary required for common diplomatic usage; a fourth letter was added for plurals, participles and genitives. When encoding the plural, genitive, or participle of a 2-letter word, the third letter would be placed apart in order to avoid confusion. Code symbols were prepared for principal countries and cities in the world, for states, major cities, and territories of the United States, and for proper names of men in English. A cipher table was to be used for those words not on the list. The first 74 pages of the code was the encode section, and contained the words in alphabetical order together with the code symbols; for example the very first word was Aaron with the symbol ABA, the last word of the first page was Acknowledge with a symbol of EA. The decode section (3-letter symbols) was not published in one sequential alphabet and was time consuming. Transmission of the code by cable was awkward because number of characters was not standard. It was not until 1876 that the 5 digit form became standard in the American ciphers. This code became the secret communication mask for American ministers in foreign legations in the years to 1876. Table 20-16 is a chart of the number of encoded lines sent from American ministers in seven major nations using this code.

Table 20-15
1/3 Sample page WE029

ekf Lamentation
elf Language
emf Languid
enf Languidly
eof Languish
epf Languishing
eqr Lapse
erf Large
esf Largely
etf Lasting
euf Lastly
evf Late
exf Latent
eyf Latently
ezf Latin
faf Latitude
fbf Later
fcf Laugh
fdf Launch
fef Lavish
fff Lavishly
fgf Lawyer
fhf Lawful
fif Lawfully
fjf Lawfulness
fkf Lawless
flf Lawlessly
fmf Lawlessness
fnf Lax
fof Laxity
fpf Laxly
fqf Laxness
frf Lay
fsf Laziness
ftf Lazy
fuf Leader
fvf League
fxf Leak
fzf Lean
gaf Leap
gbf Learning
gcf Leave
gdf Lecture
gef Lecturer
gff Left
ggf Legal
ghf Legally
gif Legibility

Table 20-16

	Russia	Netherlands	Great Britain	Mexico
1866			11	
1867				
1868			38	
1869			122	
1870	6		184	
1871	259		61	
1872	3		189	
1873			1	
1874	17			
1875	20			
1876				
Total	305		606	

	France	Spain	Germany
1866	33		
1867			
1868		7	
1869		26	
1870	27	52	11
1871	5		40
1872		31	10
1873	1	34	6
1874		20	2
1875	25		46
1876	13		
Total	71	170	115