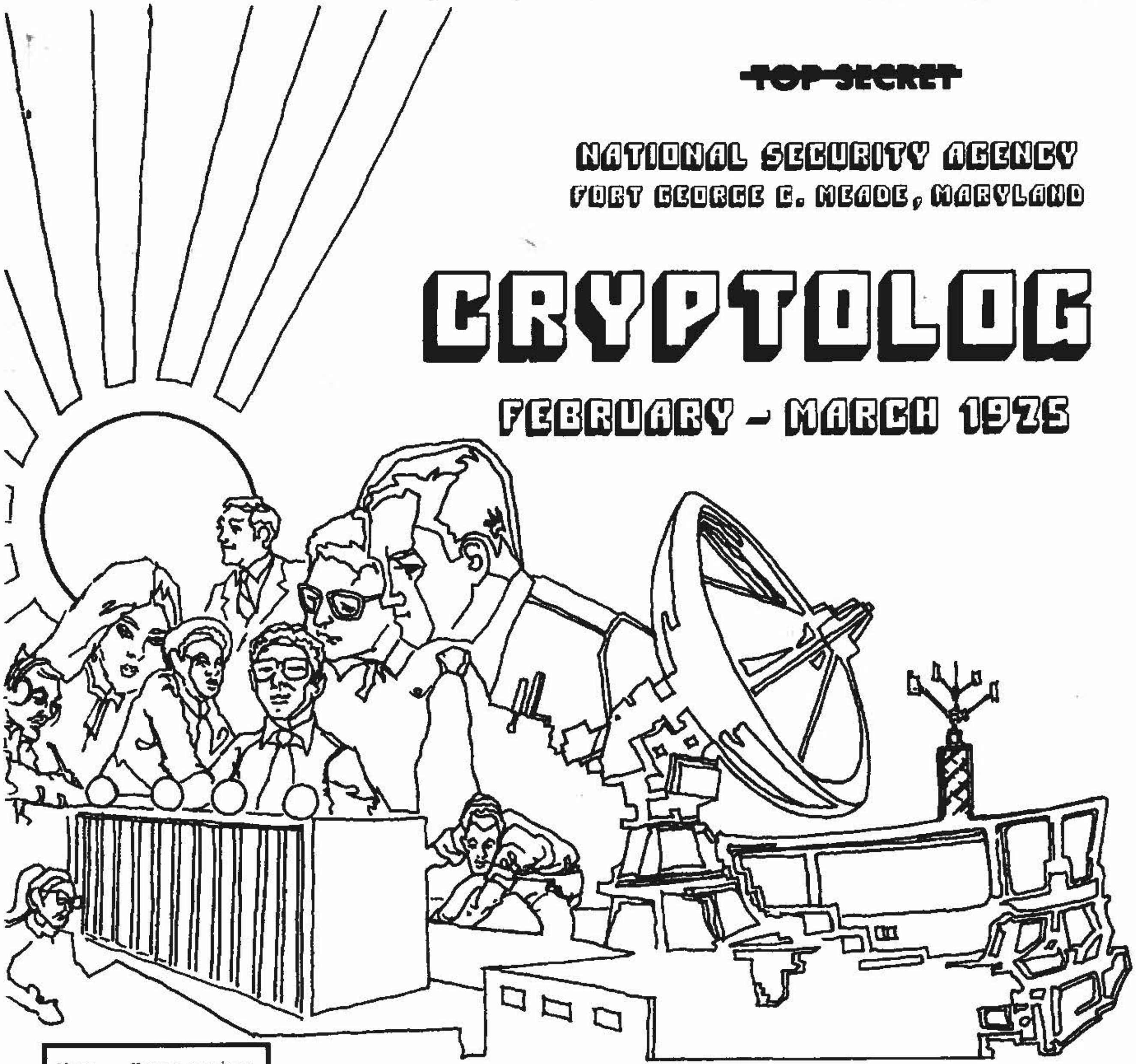
~~TOP SECRET~~

NATIONAL SECURITY AGENCY
FORT GEORGE G. MEADE, MARYLAND

CRYPTOLOG

FEBRUARY - MARCH 1975



Non - Responsive

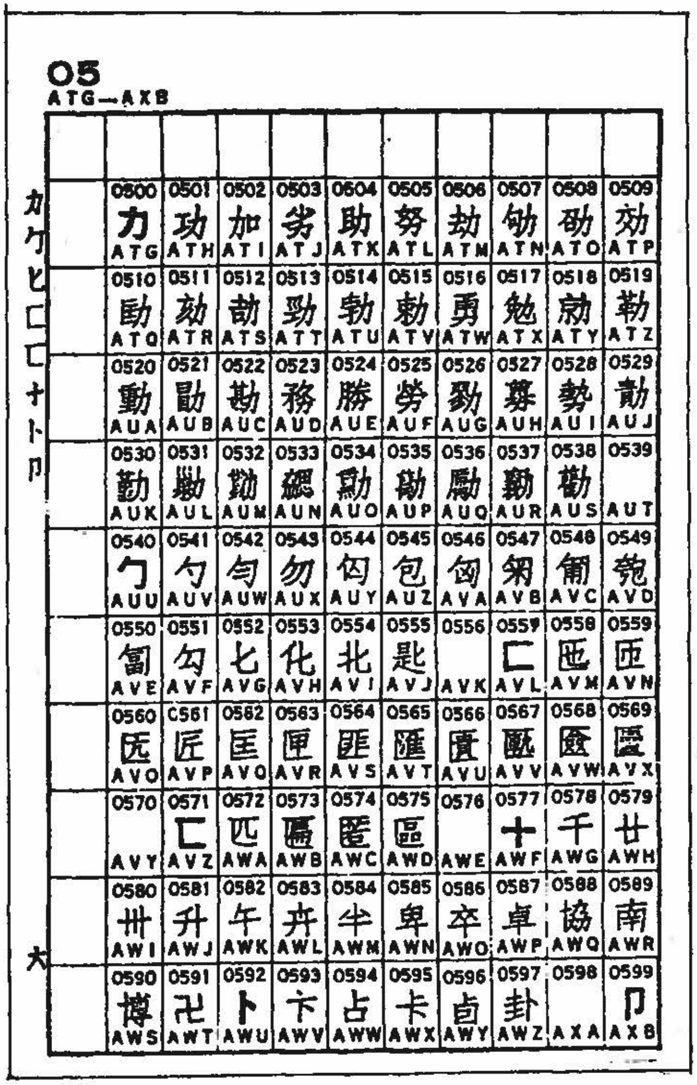
~~THIS DOCUMENT CONTAINS CODEWORD MATERIAL~~

~~TOP SECRET~~

~~Classified by DIRNSA (NSAM 123-2)~~
~~Exempt from GDS, EO 11652, Category 2~~
~~Declassify Upon Notification by the Originator~~

BASIC PATTERNS OF CHINESE CODES AND CIPHERS

William T. Mau, B4



05
ATG—AXB

0500 力 ATG	0501 功 ATH	0502 加 ATI	0503 劣 ATJ	0504 助 ATK	0505 努 ATL	0506 劫 ATM	0507 劬 ATN	0508 劭 ATO	0509 効 ATP
0510 劻 ATQ	0511 劾 ATR	0512 劼 ATS	0513 勁 ATT	0514 勃 ATU	0515 勅 ATV	0516 勇 ATW	0517 勉 ATX	0518 勍 ATY	0519 勒 ATZ
0520 動 AUA	0521 勖 AUB	0522 勘 AUC	0523 務 AUD	0524 勝 AUE	0525 勞 AUF	0526 勠 AUG	0527 募 AUH	0528 勢 AUI	0529 勣 AUJ
0530 勤 AUK	0531 勦 AUL	0532 勱 AUM	0533 勰 AUN	0534 勳 AUO	0535 勴 AUP	0536 勵 AUQ	0537 勸 AUR	0538 勷 AUS	0539 AUT
0540 勹 AUU	0541 勺 AUV	0542 勻 AUW	0543 勿 AUX	0544 匈 AUY	0545 包 AUZ	0546 匋 AVA	0547 匊 AVB	0548 匍 AVC	0549 匏 AVD
0550 匐 AVE	0551 勾 AVF	0552 匕 AVG	0553 化 AVH	0554 北 AVI	0555 匙 AVJ	0556 AVK	0557 匚 AVL	0558 匜 AVM	0559 匝 AVN
0560 匟 AVO	C561 匠 AVP	0562 匡 AVQ	0563 匣 AVR	0564 匪 AVS	0565 匯 AVT	0566 匱 AVU	0567 匭 AVV	0568 匳 AVW	0569 匵 AVX
0570 AVY	0571 匸 AVZ	0572 匹 AWA	0573 匾 AWB	0574 匿 AWC	0575 區 AWD	0576 AWE	0577 十 AWF	0578 千 AWG	0579 廿 AWH
0580 卅 AWI	0581 升 AWJ	0582 午 AWK	0583 卉 AWL	0584 半 AWM	0585 卑 AWN	0586 卒 AWO	0587 卓 AWP	0588 協 AWQ	0589 南 AWR
0590 博 AWS	0591 卍 AWT	0592 卜 AWU	0593 卞 AWV	0594 占 AWW	0595 卡 AWX	0596 卣 AWY	0597 卦 AWZ	0598 AXA	0599 卩 AXB

力 勹 匕 匚 匸 十 卜 卩 大

When is a code not a code? When it's Chinese plain text. As many, but not all, readers know, in ordinary communications the Chinese use a four-digit code to represent the thousands of characters of their written language. That code is the Standard Telegraphic Code (STC), a page of which is shown at left. Since Chinese characters cannot be sent by telegraph, this set of digital equivalents was set up to make telecommunications possible. (The four-digit groups are what we usually see in traffic; the trigraphs below the characters are an alternate set of equivalents, rarely used.)

Some version of this code has been in use in China ever since telegraphy was introduced there. The Chinese Nationalists use the older "CTC" (Chinese Telegraphic Code, also known as the "Ming Code" from the Chinese words on the cover of the book: MING MA, plain code), which contains older forms of many characters, and which reads from right to left rather than left to right.

The STC book consists of 100 pages, on each of which is a 10 x 10 matrix with single-digit row and column coordinates. In each cell is one Chinese character. The characters are arranged in "radical/stroke" order (the most common dictionary order in Chinese use), with, unfortunately, some out-of-order exceptions. The basic characters occupy the first 79 pages; the remaining pages contain short forms, additional characters, and special tables, including dates, times, and marks of punctuation, and the Latin alphabet.

The root or base of a Chinese character is the "radical." There are 214 radicals, most of which are characters by themselves: i.e., they have a meaning. But their meaning is extended or modified by the addition of one or more strokes to form another character. The number of radicals was established about 1660, and their dictionary order is fixed by the number of strokes it takes to write each one--from a single stroke (一) to 17 (龠). In STC each radical appears as a "section heading," and the characters which follow it are arranged basically in ascending number of strokes added to the basic root. (Compare this with the English alphabet with its 26 letters, only two of which have meaning when they stand alone.)

This article examines certain features of Chinese-language cryptographic code systems. Each type is treated separately, but in actual practice a code system often involves two or more of the methods of encryption discussed below. In addition, two or more methods for encoding values not in the vocabulary can be used in a single code system.

BASIC CODE PATTERNS

As we all learned in early training, codes can be one-part or two-part. One-part codes are so formatted that one book suffices for both encoding and decoding. In two-part codes the order of plain and code equivalents is so mixed that two books are required, one for encoding and one for decoding. A modified one-part code is one in which the regular pattern has been complicated in some way. Most Chinese codes are one-part.

But one of the most important questions facing the codebreaker as he looks at a new Chinese code is this:

Assuming that it is one-part, and that the values therefore occur in logical order, rather than scrambled, what is that logical order? In European languages it would be some form of alphabetic order, but Chinese has at least five possible ways of arranging a one-part code:

Radical/Stroke Order

The codes used by the People's Republic of China (PRC) may be arranged in the usual radical/stroke order which parallels the STC book, but other systematic orders--even other radical/stroke systems--are possible. Compounds can be inserted between single characters. For example, a given row of ten values in the STC book is 0100, 0101, 0102, 0103, etc., through 0109. The corresponding values in a radical/stroke order code might be 0100, 0100 0226, 0101, 0102, 0102 7022, etc.

Phonetic Order

An alphabetic order for Chinese-language values can be achieved by one of several systems of phonetic representation of characters. Strangely, the system usually seen is not the Pinyin, introduced in 1958 (with which we at NSA prefer to work for convenience' sake), but the older, National Phonetic, system, designed in the early 1920's in imitation of the Japanese *kana*. STC books published in China contain these 37 phonetic symbols in the otherwise blank cells 9720--9756. Thus while the symbols in the names of Mao Ze Dong, Shanghai and Beijing (Peking/Peiching) would occur in the same order in a Pinyin listing as in English alphabetical order: Bei, Dong, Hai, Jing, Shang, Ze, under the National Phonetic system they would be listed in the order: Bei, Mao, Dong, Hai, Shang, Ze. (This system is known, from the first four syllables in it, as the Bo-Po-Mo-Fo system.)

Total Strokes Order

The same names might be ordered by the number of strokes with which they are written, using the number of strokes in the new (circa 1958) short forms where applicable. Thus arranged, they would read: Shang (上 3 strokes), Mao (毛 4 strokes), Bei (北 5 strokes), Dong (东 5 strokes), Jing (京 8 strokes), Ze (泽 8 strokes), Hai (海 10 strokes). Characters with the same number of strokes may be arranged within that category by either phonetic or radical order.

Sentential/Category Order

Many PRC codes, especially military codes, are in the form of charts rather than books. Dimensions vary, but the 9 x 9 matrix is most common. The usual number of matrices in a code of this type is from six to nine.

Such charts frequently use popular phrases and sentences to fill in the rows of the matrices. Categories such as time (year, month, day, hour) and points of the compass (east, south, west, north) are listed in adjoining cells in other matrices. (Of the characters used in our names example, above, "Dong," the literal meaning of which is east, and "Bei," literally north, would be found with south and west.) The sentential/category code is often referred to as "modified one-part" because it is "one-part" (requires only one book) to has someone who has memorized the sentences and has the chart in front of him.

COMPARATIVE ORDER OF THE CHARACTERS IN THE NAMES "Mao Ze Dong," "Shanghai," "Beijing"

STC and Radical/ Stroke	Pin-yin	Nat'l Phon.	Total Stroke	Sentential
SHANG (0006) 上	BEI	BEI	SHANG (3)	MAO
JING (0079) 京	DONG	MAO	MAO (4)	ZE
BEI (0554) 北	HAI	DONG	BEI (5)	DONG
DONG (2639) 东	JING	HAI	DONG (5)	SHANG
MAO (3029) 毛	MAO	JING	JING (8)	HAI
HAI (3189) 海	SHANG	SHANG	ZE (8)	BEI
ZE (3419) 泽	ZE	ZE	HAI (10)	JING

EXAMPLE OF SENTENTIAL ARRANGEMENT

	1		2		3		4		5		6	
10	各	部	注	意	隱	蔽	裝	備	加	強	防	空
11			前	進	受	阻	請	求	砲	火	支	援
12	情	況	突	然	變	化	人	員	傷	亡	嚴	重
13	命	令	你	部	立	即	作	好	戰	斗	準	備
14												

Line 10: All units are to pay attention to concealing their equipment and strengthening air defense. Do not expose targets.

Line 11: Our advance is blocked. Request artillery support to help in completing our mission.

Line 12: The situation has suddenly changed. Casualties are heavy.

Line 13: Your unit is directed to complete combat preparations at once.

	114	132	105	116
ENCODED TEXT:	請求	你部	加強	支援

Request your-unit increase support.

PROVISION FOR ADDITIONAL VALUES

In practice, the reconstruction of codes is facilitated by the fact that the code vocabulary is not all-inclusive. Some of the characters in a given message may not be in the code vocabulary at all, and the code clerk must have some means of encoding the absent characters. There are several ways to deal with "missing" characters, and each makes the bookbreaker's task a bit easier. Among them are the phonetic spell table, phonetic variation, character construction and dissection, and enciphered STC.

Spell Tables

Chinese codes often contain a subsystem for "spelling out" characters which do not occur in the code vocabulary. Most common are phonetic values or substitution tables for STC monomes or dinomes. Sometimes the code groups for digital values are used, with a flag group, to "spell" the digits of an STC group. A device used in pre-Communist codes was a table of radicals to be added to the character represented by the preceding code group. Since the spell tables can represent the Chinese language fully, they give the effect of a cipher within a code.

Phonetic Representation

A chart containing all the initials and finals that compose National Phonetic representations of Chinese characters permits the user to spell the sound of a missing character. Reconstruction of a chart such as the one which follows greatly assists the bookbreaker's recovery of values in the basic system. (For our own convenience, we at NSA write the cell contents in Pinyin instead of using the actual phonetic symbols.)

	10	20	30	40	50	60	70	80		
10		B	K	Z	ai	uo	ian	ung		
20		P	H	C	ei	uai	in			
30		M	J	S	ao	ui	iang			
40		F	Q	A	ue	ou	ing			
50		D	X	O	ia	an	iong			
60		T	ZH	E	ie	en	uan			
70		N	CH	I	iao	ang	un			
80		L	SH	U	iu	eng	uang			
90		G	R	Ü	ua	ong	er			

Another way to use sound rather than meaning is phonetic variation. A special flag group indicates that the group which follows is not to be read as its true value but that the character intended is a homonym of the plaintext equivalent of the group sent. For example, in the phrase 四分 8490 中 , the flag group 8490 indicates that the next group 中 (Zhong) is being used instead of the intended character 鈡 , which is also pronounced "Zhong" but which is not in the code vocabulary. Often one digit in the flag group will change to reflect a specific tone among the the four tones that the intended character has. The flag thus more exactly identifies the intended character among its homonyms.

Character Construction and Dissection

Using flag groups to select only a part of the preceding character (the part itself being a separate character) is known as character dissection. Character construction, on the other hand, involves flag groups that instruct the recipient to "take part A of character X and add it to part B of character Y to create the intended character." An example of dissection might break out to "Take the right-hand side of HAI (海)," which gives us MEI (每). An example of character construction would state: "Take the left-hand side of HAI (海) and add it to the character JING (京)," which gives us LIANG (涼).

Enciphered STC

Flags might also indicate the presence of STC groups within the text of coded messages. Such a flag would indicate that the group which follows is not a value in the code but is the STC group representing the character intended. This would be equivalent to inserting plaintext, so the STC group is often enciphered by an additive or by transposition.

CIPHER SYSTEMS

There have been some substitution ciphers which assigned cipher equivalents to phonetic values, but enciphered Chinese plain text, in NSA parlance, is usually some encipherment of STC.

Additive, substitution, and transposition are all used. An additive cipher may be as simple as adding a four-digit constant stutter to each group, turning 6153 0132 0932 0171 (請 你 回 來) into 7264 1243 1043 1282 by the addition of 1111; it may be as secure as a one-time running key in which a different four-digit group is added to each plain STC group; or it may be some intermediate method.

Another Chinese cipher not infrequently encountered is repaginated STC. The repagination may be merely an end-around shift of the page-number sequence or it may involve a random scramble of the page numbers. Either shift or scramble may extend to the coordinates of row or column, or both, on each page.

Local transposition within a group or the insertion of nulls can disguise the basic STC group. Thus with transposition *cabd*, 6153 0132 0932 0171 (請 你 回 來) becomes 5613 3012 3092 7011. Nulls can stretch each group to five digits. Inserting a 0 between *b* and *c* in the groups of our example gives us 61053 01032 09032 01071. In practice, transposition and insertion of nulls are often combined, so that, in the example given, the original message would become 56013 30012 30092 70011.

While it might be argued that the additive and repaginations belong under "codes" rather than "ciphers," we include them under ciphers because the same basic extensive vocabulary is used. Chinese codes selectively restrict this basic vocabulary.

Systems such as those discussed here have been used by the PRC in its communications. Knowing some of these methods certainly lightens the work of the bookbreaker. As another writer has said, "Chinese codebooks are nothing more than a compilation of written characters, expressed numerically for the purpose of telegraphic communications. The real clue to the structure of a code lies in the arrangement of these characters."

~~(TOP SECRET UMBRA)~~

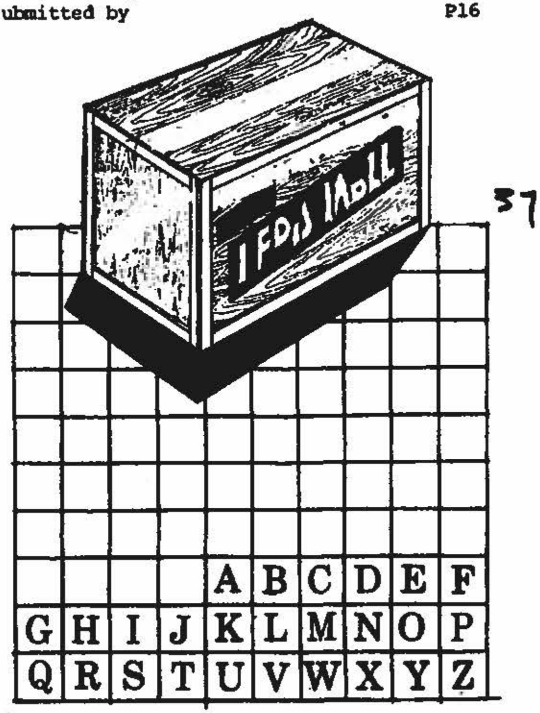
[redacted]

[redacted]

Non - Responsive

Can you make out the name?

A real-life puzzle submitted by GLENN EMERY P16



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

We know that in a certain non-English codebook the roman alphabet occurs as shown on a code page consisting of a 10 x 10 matrix (see above). The initial dinome of a code group represents the page, the final dinome column and row, respectively.

A message is received which contains the following groups in mid-text:

(3742 3792 3732 3767) (3742 3767 3732 3775 3772)

Parentheses have been observed in other messages setting off groups which represent special categories of information supplemental to the main body of the code, and context indicates that the parenthetical groups in this message represent the name of a powdered milk available in the Southeast Asian market. So it is suspected that the groups are two words using the roman alphabet, whose page has been renumbered as 37.

What are the words, and which column and row coordinates can be recovered?

(Solution next month.)

~~(TOP SECRET UMBRA)~~