

Chapter 4

How to Solve a Problem in *The Cryptogram*

A typical issue of *The Cryptogram* contains 13-14 pages of cipher puzzles, or "cons" (constructions). No information is included on how to solve these problems; all you are given is the type of cipher, a hint as to the subject of the con, and often a crib (tip) which is a piece of the plaintext. If you are new to decipherment, you will probably need some help on how to proceed. Here are a few notes; more details can be found in the literature.

Enciphered Tips

In real cryptography one would have an indication of the subject matter, and probably a wealth of ciphertext. Space limits us, so we try to compensate by giving titles and tips. Enciphered tips are those that are optional for solving. They are enciphered so that those who want a greater challenge can ignore them. For those who need the help, the tips are recovered by running down (or up) the alphabet, as tips are always enciphered in a Caesar cipher. The Caesar shift is not normally more than six in either direction.

EXAMPLE 1:

Enciphered tip: Z L Z H I B
 A M A I J C
 B N B J K D
Plain tip: C O C K L E

EXAMPLE 2:

Enciphered tip: U Q E K G V A
 T P D J F U Z
Plain tip: S O C I E T Y

Here the Caesar shifts are -3 and +2. Unless the tip is simply giving the period of the cipher, it is ALWAYS found somewhere in the message. Enciphered tips are always in UPPERCASE, and plain tips are in lowercase.

Keyword Recovery

A keyword is an easily remembered group of words, letters, or numbers originally, and still used, by correspondents who wish to use known methods of encipherment and still maintain secrecy. Many ciphers use a keyword to generate ciphertext alphabets to scramble the order of the message before, during or after encipherment. Recovery of the keyword offers an alternative or parallel path to solution alongside direct recovery of the plaintext. As the plaintext is revealed, so is the keyword, and guessing a letter in one may give a letter in the other that might aid solution.

Keywords are discussed in more detail in Chapter 8.

General Properties of Letters

A large vocabulary is helpful, but even more useful is a knowledge of the habits of letters and their relationships to each other in English (or the language of the message). The Frequency Table for English is:

E	T	A	O	N	I	R	S	H	L	D	C	U	P	F	M	W	B	G	V	K	Q	X	J	Z			
13	9	8		7		6		4		3		2		1		-											total: 100

The high-frequency letters ETAONIRSH make up about 70% of plain text. Vowels AEIOU and Y make up about 40% of the text. Consonants LNRST make up about 35% of the text. The low frequency letters occupy less than 3% of the text but are important because of their rarity and because adjacent letters (contacts) are more likely to be vowels than consonants.

We will now take a look at the various letters in more detail. Do not be intimidated by these statistics! Refer to them as you solve; be aware that they exist; and over time you will find they become second nature to you.

Vowels

1-letter words	:	A, I; O occasionally.
2-letter words	:	begin with A, I, O, U; end with E, O, Y.
Doubles	:	O, E often double; A, U, I, Y rarely double.
Digraphs	A	: follows E, O (EA is most frequent); reverses with I, U, Y.
	E	: precedes A; follows O; reverses with I, U, Y.
	I	: follows U, Y; reverses with A, E, O.
	O	: precedes A, U; reverses with I, Y.
	U	: follows O; precedes I; reverses with A, E.
	Y	: follows U; precedes I; reverses with A, E, O.

Common Positions:

A	:	initial and 2nd from end.
E	:	2nd and final, also scattered throughout.
I	:	3rd from end.
O	:	2nd and final.
U	:	initial and 2nd from end.
Y	:	final.

Vowel-Consonant Digraphs

ER-RE is the most frequent reversed digraph.

A:	follows H;	precedes N,T,S;	reverses with R.
E:	follows H;	precedes S,D;	reverses with R,T.
I:	follows H,R,D;	precedes N;	reverses with T,S.
O:	follows T,S,H;	precedes N;	reverses with R,L.
U:	follows S,T,F;	precedes N;	reverses with P,B.
Y:	follows L,R,T,N;	precedes S.	

High-Frequency Consonants HLNRST

- H: likes 1st, 2nd, last position; precedes vowels; follows W, S, C, T. Note TH and GHT.
- L: likes 2nd, next-to-last; prefers vowel contacts; follows P, C, B; precedes P, D; doubles at 3-4 in 6-letter words; before final S, Y in 5-letter words; in last position in 4-letter words.
- N: likes last and next-to-last position; follows vowels; precedes D, T, G, S, C.
- R: likes 2nd and next-to-last (thus looks like a vowel, BUT it reverses freely with vowels); follows B, P, T; precedes T, S; doubles freely; often reverses with T (a common consonant reversal); seldom follows S.
- S: likes last (very strongly), 1st and third position; doubles freely at middle and last positions; follows vowels and D, T, R, N; precedes vowels and T, H, P, C, M; reverses often with T.
- T: likes 1st, last and next-to-last (also scatters); doubles freely; follows vowels and B, C, F, L, N, P, R, S, X; precedes vowels and H, R.

The following two consonants (with E, S, T, N, R, Y) end most English words:

- D: likes 1st and last position; prefers vowel contacts; doubles freely when followed by L; follows L, N, R; precedes R, W, L.
- G: likes last (strongly), sometimes 1st, 3rd from last and next-to-last; doubles freely when followed by L; follows vowels and D, R, N; precedes vowels and H, R, L.

Other Consonants

(These, with T, O, S, begin most English words).

- B follows vowels; precedes vowels, L, R; doubles when followed by L.
- C follows vowels, S, N; precedes vowels, H, T, L, R, K; doubles.
- F follows vowels; precedes vowels, T, R; doubles within and last.
- J usually initial only, precedes O, U; never doubles.
- M follows vowels, S, R; precedes vowels, P; doubles within.
- P follows vowels, R, L, M, S; precedes vowels, R, L, T; doubles freely.
- V follows vowels, L, R; precedes vowels.
- W follows vowels, D, S, T; precedes vowels, H, R.

Low-Frequency Consonants

- K likes first position followed by vowel or N; last position preceded by N, R, C, L; otherwise contacts vowels.
- Q likes 1st, 2nd, 3rd positions; normally followed by U; preceded by E, O, N, S, C.
- X follows vowels and N; precedes vowels and C, H, P, T.
- Z contacts vowels on both sides normally. (NB: UK uses "S" for USA "Z" in many instances.)

The Index of Coincidence

The Index of Coincidence (IC) is a powerful test to help the analyst decide whether a set of ciphertext has been encrypted with the same alphabet. It measures the roughness of the frequencies: that is, if the letter frequencies are about equal, the frequencies are smooth. Natural language tends to have a rough frequency distribution. To calculate the IC, take a frequency count of the individual letters. For each letter, multiply the frequency by the frequency minus one, then add them all together. Divide the sum by $N * (N-1)$, where N is the number of letters in the sample. The result is the IC. For example, the frequencies of the first sentence in this paragraph are:

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
8	2	7	5	22	3	0	10	8	0	0	4	1	6	5	5	0	4	6	14	1	0	3	2	2	0

The numerator is $8*7 + 2*1 + 7*6 + 5*4 + \dots + 2*1 + 1*(0) + 0*(-1) = 1050$, and the denominator is $118 * 117 = 13,806$, so the IC is 0.0761. The expected IC for English is about 0.066, but in practice, because of statistical variations in the sample, anything over 0.055 is quite likely to be from a single alphabet. If the value is near 0.0385, the expected value for random text, the sample is almost certainly not from a single alphabet.

Different languages will have different expected IC's.

Solving Aristocrats

An Aristocrat is a simple substitution cipher with word divisions. No letter stands for itself.

Check the title first, and think of any word that might appear in the text. Look for short common words and pattern words. Make a frequency count to determine likely letter equivalencies. Look for 3 and 4-letter words, especially those containing TH. Look for common endings and beginnings.

Look for these short words: an, in, is, it, on, to, and, are, has, his, her, not, see, the, was, why, you, from, into, once, have, that, than, this, there, these, those.

Look for these word beginnings: an, at, be, de, dr, en, in, no, pre, pro, re, se, th, un.

Look for these word endings: ance, ant, ate, (a)ble, ded, ed, en, er, ere, es, ese, ess, est, ful, ght, ine, ing, ion, is, ist, ive, ll, lly, ment, ous, rst, ses, sts, tion.

Computer solution often involves the use of large word lists to provide word-matches, which by trial and error can then lead to a solution.

Solving Xenocrypts

A Xenocrypt may be any of the regular cipher types, but using a foreign language. It is usually not necessary to know the language to solve the ciphers, especially when tips are given. However, such things as small dictionaries or word lists, beginner's books, and the familiarity gained by previous attempts can be very useful.

Xenocrypts are attacked in the same way as problems in English, but using, of course, frequency tables and other data for the language in question. Additional data can be found in the literature; frequency tables for some languages are given below. Other references can be found in *The Cryptogram*, *Elcy*, and *Xenocrypt Handbook*.

Frequency Tables

References

English:	ETAONIRSHLDCUPFMWYBGVKQXJZ	
Dutch:	ENIATORDLSGKHVUWBJMPZCFYXQ	<i>Xeno Handbook</i>
Esperanto:	AIEONLRSTKJUDMPVBGFZCH	<i>Xeno Handbook</i>
French:	EANRSITUOLDCMPVBFHQJZXY	MA86, <i>Elcy</i>
German:	ENIRSADTUGHOLBMCWFKVZPJQXY	MA89, <i>Elcy</i>
Interlingua:	EAILNOSTRUDCMPVGBFHQJWYZK	MJ75
Italian:	EAIOLNRTSCDMPUVGZFBHQ	MJ86, <i>Elcy</i>
Latin:	IEUTAMSNRODLVCPQBFHXHJKWYZ	ND50
Portuguese:	AEORSINDMTUCLPQVFGHBJZX	<i>Xeno Handbook</i>
Spanish:	EAOSRNIDLCTUMPGYBQVHFZ	JF86, <i>Elcy</i>
Swedish:	AENRTSIOMGKLDVFBCHPUYJXQWZ	JA81

Note that these figures are statistical averages; thus the orders in *Elcy* are somewhat different from those in *The Cryptogram* and *Xeno Handbook* references.

Tips often give the identity of letters appearing only once, singletons, or not at all. Letters not appearing in the plaintext alphabet are marked with an asterisk.

The Cipher Exchange

This department of *The Cryptogram* contains a selection of ciphers which do not use simple substitution. Some 60 ciphers are in current use by the ACA, and these are detailed in Chapter 9. Methods of solution are found in the literature and are continually being improved, particularly with the application of computers. Exchanging ideas with other member may also help.

The Analyst Corner

These ciphers are considered somewhat harder and may break the "rules" for the cipher in some way, such as by the omission of a tip. The ciphers in this department may be longer than those in the "Cipher Exchange" to facilitate statistical analysis. A generous title or short narrative to "set the stage" is sometimes included.

Ornamentals

Ornamentals are pictorial ciphers, often found as front cover designs for *The Cryptogram* (See Steganography). The ciphers are usually Aristocrats or Patristocrats, but don't rule out other cipher types. An ornamental should be treated as any Unknown. The challenge is to determine how the artist has hidden the message in the pattern, and then to solve it. Look for a typical "cell" that could represent a letter. Checking the dimensions of the figure may give a clue for there are likely to be 75-100 characters hidden in it.

Specials and Challenges

These are ciphers that the Editors consider more difficult than normal (such as a Patristocrat without plaintext "e"s), or examples of new types introduced in articles. Because of the nature of these ciphers, they are not included in the requirements for a complete (*) solution in the Solvers List, although they do count in the total solved.