CLASSICAL CRYPTOGRAPHY COURSE
BY LANAKI
September 27, 1995

LECTURE 1
SIMPLE SUBSTITUTION

INTRODUCTION

Cryptography is the science of writing messages that no one except the intended receiver can read. Cryptanalysis is the science of reading them anyway. "Crypto" comes from the Greek 'krypte' meaning hidden or vault and "Graphy" comes from the Greek 'grafik' meaning writing. The words, characters or letters of the original intelligible message constitute the Plain Text (PT). The words, characters or letters of the secret form of the message are called Cipher Text (CT) and together constitute a Cryptogram.

Cryptograms are roughly divided into Ciphers and Codes.

William F. Friedman defines a Cipher message as one produced by applying a method of cryptography to the individual letters of the plain text taken either singly or in groups of constant length. Practically every cipher message is the result of the joint application of a General System (or Algorithm) or method of treatment, which is invariable and a Specific Key which is variable, at the will of the correspondents and controls the exact steps followed under the general system. It is assumed that the general system is known by the correspondents and the cryptanalyst. [FRE1]

A Code message is a cryptogram which has been produced by using a code book consisting of arbitrary combinations of letters, entire words, figures substituted for words, partial words, phrases, of PT. Whereas a cipher system acts upon individual letters or definite groups taken as units, a code deals with entire words or phrases or even sentences taken as units. We will look at both types of systems in this course.

The process of converting PT into CT is Encipherment. The reverse process of reducing CT into PT is Decipherment.

Cipher systems are divided into two classes: substitution and transposition. A Substitution cipher is a cryptogram in which the original letters of the plain text, taken either singly or in groups of constant length, have been replaced by other letters, figures, signs, or combination of them in accordance with a definite system and key. A Transposition cipher is a cryptogram in which the original letters of the plain text have merely been rearranged according to a definite system. Modern cipher systems use both substitution and transposition to create secret messages.

SUBSTITUTION AND TRANSPOSITION CIPHERS COMPARED

The fundamental difference between substitution and transposition methods is that in the former the normal or conventional values of the letters of the PT are changed, without any change in the relative positions of the letters in their original sequences, whereas in the latter only the relative positions of the letters of the PT in the original sequences are changed, without any changes to the conventional values for the letters. Since the methods of encipherment are radically different in the two cases, the principles involved in the cryptanalyses of both types of ciphers are fundamentally different. We will look at the methods for determine whether a cipher has been enciphered by substitution or transposition.

SIMPLE SUBSTITUTION

Probably the most popular amateur cipher is the simple substitution cipher. We see them in newspapers. Kids use them to fool teachers, lovers send them to each for special meetings, they have been used by the Masons, secret Greek societies and by fraternal organizations. Current gangs in the Southwest use them to do drug deals. They are found in literature like the *Gold Bug* by Edgar Allen Poe, and death threats by the infamous Zodiak killer in San Francisco in the late 1960's.

The Aristocrats (A1-A25) in the Aristocrats Column of "The Cryptogram" are all simple substitution ciphers in English. Each English plain text letter in all its occurrences in the message is replaced by a unique English ciphertext letter. The

mathematical process is called one-to-one contour mapping. It is unethical (and a possible wedge for the analyst) to use the same ciphertext letter for substitution for a plaintext letter.

A recurring theme of my lectures is that all substitution ciphers have a common basis in mathematics and probability theory. The basis language of the cipher doesn't matter as long as it can be characterized mathematically. Mathematics is the common link for deciphering any language substitution cipher. Based on mathematical principles, we can identify the language of the cryptogram and the break open its contents.

FOUR BASIC OPERATIONS OF CRYPTANALYSIS

William F. Friedman presents the fundamental operations for the solution of practically every cryptogram:

(1) The determination of the language employed in the plain text version.

(2) The determination of the general system of cryptography employed.

(3) The reconstruction of the specific key in the case of a cipher system, or the reconstruction of, partial or complete, of the code book, in the case of a code system or both in the case of an enciphered code system.

(4) The reconstruction or establishment of the plain text.

In some cases, step (2) may proceed step (1). This is the
classical approach to cryptanalysis. It may be further reduced to:

1. Arrangement and rearrangement of data to disclose non-random characteristics or manifestations (i.e. frequency counts, repetitions, patterns, symmetrical phenomena)

2. Recognition of the nonrandom characteristics or manifestations when disclosed (via statistics or other techniques)

3. Explanation of nonrandom characteristics when recognized. (by luck, intelligence, or perseverance)

Much of the work is in determining the general system. In the final analysis, the solution of every cryptogram involving a form of substitution depends upon its reduction to mono-alphabetic terms, if it is not originally in those terms[FRE1]

OUTLINE OF CIPHER SOLUTION

According to the Navy Department OP-20-G Course in Cryptanalysis, the solution of a substitution cipher generally progresses through the following stages:

(a) Analysis of the cryptogram(s)

(1) Preparation of a frequency table.
(2) Search for repetitions.
(3) Determination of the type of system used.
(4) Preparation of a work sheet.
(5) Preparation of individual alphabets (if more than one)
(6) Tabulation of long repetitions and peculiar letter distributions.

(b) Classification of vowels and consonants by a study of:

(1) Frequencies
(2) Spacing
(3) Letter combinations
(4) Repetitions

(c) Identification of letters.

(1) Breaking in or wedge process
(2) Verification of assumptions.
(3) Filling in good values throughout messages
(4) Recovery of new values to complete the solution.

(d) Reconstruction of the system.

(1) Rebuilding the enciphering table.
(2) Recovery of the key(s) used in the operation of the system
(3) Recovery of the key or keyword(s) used to construct the alphabet sequences.

All steps above to be done with orderly reasoning. It is not an exact mechanical process. [OP20]

Since this is a course in Cryptanalysis, lets start cracking some open.

EYEBALL

While reading the newspaper you see the following cryptogram. Train your eye to look for wedges or 'ins' into the cryptogram. Assume that we dealing with English and that we have simple substitution. What do we know? Although short, there are several entries for solution. Number the words. Note that it is a quotation (12, 13 words with * represent a proper name in ACA lingo).

A-1. Elevated thinker. K2 (71) LANAKI

```
  1          2                    3          4     5       5         6       7

F Y V   Y Z X Y V E F   I T A M G V U X V   Z E   F A   I T A M   F Y Q F   M V

  8             9           10         11              12

Q D V   E J D D A J T U V U   R O   H O E F V D O.   * Q G R V D F

          13

* E S Y M V Z F P V D
```

ANALYSIS OF A-1.

Note words 1 and 6 could be: ' The....That' and words 3 and 5 use the same 4 letters I T A M . Note that there is a flow to this cryptogram The _ _ is? _ _ and? _ _. Titles either help or should be ignored as red herrings. Elevated might mean "high" and the thinker could be the proper person. We also could attack this cipher using pattern words (lists of words with repeated letters put into thesaurus form and referenced by pattern and word length) for words 2, 3, 6, 9, and 11.

Filling in the cryptogram using [The... That] assumption we have:

```
  1          2                    3          4     5       5         6       7
t h e   h     h e s t             e     e     s   t              t h a t     e
F Y V   Y Z X Y V E F   I T A M G V U X V   Z E   F A   I T A M   F Y Q F   M V

  8             9           10         11              12
a   e                   e                t e          a     e   t
Q D V   E J D D A J T U V U   R O   H O E F V D O.   * Q G R V D F

          13
      h   e   t   e
* E S Y M V Z F P V D
```

Not bad for a start. We find the ending e_t might be 'est'. A two letter word starting with t_ is 'to'. Word 8 is 'are'. So we add this part of the puzzle. Note how each wedge leads to the next wedge. Always look for confirmation that your assumptions are correct. Have an eraser ready to start back a step if necessary. Keep a tally on which letters have been placed correctly. Those that are unconfirmed guesses, signify with ? Piece by piece, we build on the opening wedge.

```
  1         2                  3          4     5       5        6      7
t h e   h     h e s t       o     e     e     s   t o       o     t h a t     e
F Y V   Y Z X Y V E F   I T A M G V U X V   Z E   F A   I T A M   F Y Q F   M V

  8             9            10         11              12
a r e   s   r r         e             s t e r       a     e r t
Q D V   E J D D A J T U V U   R O   H O E F V D O.   * Q G R V D F

          13
  s   h   e   t   e r
* E S Y M V Z F P V D
```

Now we have some bigger wedges. The s_h is a possible 'sch' from German. Word 9 could be 'surrounded.' Z = i. The name could be Albert Schweitzer. Lets try these guesses. Word 2 might be 'highest' which goes with the title.

```
  1         2                  3          4     5       5        6      7
t h e   h i g h e s t     n o w l e d g e   i s   t o     n o w   t h a t   w e
F Y V   Y Z X Y V E F   I T A M G V U X V   Z E   F A   I T A M   F Y Q F   M V

  8             9           10         11              12
a r e   s u r r o u n d e d           s t e r       a l b e r t
Q D V   E J D D A J T U V U   R O   H O E F V D O.   * Q G R V D F

          13
  s c h w e i t z e r
* E S Y M V Z F P V D
```

The final message is: The highest knowledge is to know that we are surrounded by mystery. Albert Schweitzer.

Ok that's the message, but what do we know about the keying method.

KEYING CONVENTIONS

Ciphertext alphabets are generally mixed for more security and an easy pneumonic to remember as a translation key. ACA ciphers are keyed in K1, K2, K3, K4 or K()M for mixed variety. K1 means that a keyword is used in the PT alphabet to scramble it. K2 is the most popular for CT alphabet scrambling. K3 uses the same keyword in both PT and CT alphabets, K4 uses different keywords in both PT and CT alphabets. A keyword or phrase is chosen that can easily be remembered. Duplicate letters after the first occurrence are deleted.

Following the keyword, the balance of the letters are written out in normal order. A one-to-one correspondence with the regular alphabet is maintained. A K2M mixed keyword sequence using the word METAL and key DEMOCRAT might look like this:

```
4  2  5  1  3
M  E  T  A  L
=============
D  E  M  O  C
R  A  T  B  F
G  H  I  J  K
L  N  P  Q  S
U  V  W  X  Y
Z
```

the CT alphabet would be taken off by columns and used:

CT: OBJQX EAHNV CFKSY DRGLUZ MTIPW

Going back to A-1. Since it is keyed aa a K-2, we set up the PT alphabet as a normal sequence and fill in the CT letters below it. Do you see the keyword LIGHT?

```
PT  abcdefghijklmnopqrstuvwxyz
CT  QRSUVWXYZLIGHTABCDEFJKMNOP
          ----------
KW = LIGHT
```

In tough ciphers, we use the above key recovery procedure to go back and forth between the cryptogram and keying alphabet to yield additional information.

To summarize the eyeball method:

1. Common letters appear frequently throughout the message but don't expect an exact correspondence in popularity.

2. Look for short, common words (the, and, are, that, is, to) and common endings (tion, ing, ers, ded, ted, ess,

3. Make a guess, try out the substitutions, keep track of your progress. Look for readability.

GENERAL NATURE OF ENGLISH LANGUAGE

A working knowledge of the letters, characteristics, relations with each other, and their favorite positions in words is very valuable in solving substitution ciphers.

Friedman was the first to employ the principle that English Letters are mathematically distributed in a unilateral frequency distribution:

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

That is, in each 100 letters of text, E has a frequency (or number of appearances) of about 13; T, a frequency of about 9; K Q X J Z appear so seldom, that their frequency is a low decimal.

Other important data on English (based on Hitt's Military Text):

6 Vowels: A E I O U Y	= 40 %
20 Consonants:	
5 High Frequency (D N R S T)	= 35 %
10 Medium Frequency (B C F G H L M P V W)	= 24 %
5 Low Frequency (J K Q X Z)	= 1 %
	====
	100.%

The four vowels A, E, I, O and the four consonants N, R, S, T form 2/3 of the normal English plain text. [FR1]

Friedman gives a Digraph chart taken from Parker Hitts Manual on p22 of reference. [FR2]

The most frequent English digraphs per 200 letters are:

```
TH--50      AT--25      ST--20
ER--40      EN--25      IO--18
ON--39      ES--25      LE--18
AN--38      OF--25      IS--17
RE--36      OR--25      OU--17
HE--33      NT--24      AR--16
IN--31      EA--22      AS--16
ED--30      TI--22      DE--16
ND--30      TO--22      RT--16
HA--26      IT--20      VE--16
```

The most frequent English trigraphs per 200 letters are:

THE--89	TIO--33	EDT--27
AND--54	FOR--33	TIS--25
THA--47	NDE--31	OFT--23
ENT--39	HAS--28	STH--21
ION--36	NCE--27	MEN--20

Frequency of Initial and Final Letters

Letters--	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
Initial--	9	6	6	5	2	4	2	3	3	1	1	2	4	2	10	2	-	4	5	17	2	-	7	-	3	-
Final --	1	-	1	0	17	6	4	2	-	-	1	6	1	9	4	1	-	8	9	11	1	-	1	-	8	-

Relative Frequencies of Vowels.

A 19.5% E 32.0% I 16.7% O 20.2% U 8.0% Y 3.6%

Average number of vowels per 20 letters, 8.

Becker and Piper partition the English language into 5 groups based on their Table 1.1 [STIN], [BP82]

Table 1.1
Probability Of Occurrence of 26 Letters

Letter	Probability	Letter	Probability
A	.082	N	.067
B	.015	O	.075
C	.028	P	.019
D	.043	Q	.001
E	.127	R	.060
F	.022	S	.063
G	.020	T	.091
H	.061	U	.028
I	.070	V	.010
J	.002	W	.023
K	.008	X	.001
L	.040	Y	.020
M	.024	Z	.001

Groups
1. E, having a probability of about 0.127

2. T, A, O, I, N, S, H, R, each having probabilities between 0.06 - 0.09

3. D, L, having probabilities around 0.04

4. C, U, M, W, F, G, Y, P, B, each having probabilities between 0.015 - 0.023.

5. V, K, J, X, Q, Z, each having probabilities less 0.01.

LETTER CHARACTERISTICS AND INTERACTIONS

ELCY gives Data for English, German, French, Italian, Spanish, Portuguese in her Appendices, p218 ff. She also give tables of letter contact data. [ELCY]

LANAKI published data on English and 10 different languages as well as expanded work on Chinese. It is available at the CDB. [NIC1] [NIC2]

S-TUCK gives detailed English, French and Spanish letter characteristics in her book. [TUCK]

Friedman in his Military Cryptanalytics Part I - Volume 1 gives charts showing the lower and upper limits of deviation from theoretical (random) for the number of vowels, high, low, medium frequency consonants, blanks in distributions for plain text and random text for messages of various lengths. [FR1]

Friedman in his Military Cryptanalytics Part I - Volume 2 give a veritable pot puree of statistical data on letter frequencies, digraphs, trigraphs, tetragraphs, grouped letters, relative log data, special purpose data, pattern words, idiomorphic data, standard endings, initials, foreign language data [German, French, Italian, Spanish, Portuguese and Russian], classification of systems used in concealment, nulls and literals. [FR2]

Sinkov assigns log frequencies to digraphs to aid in identification. The procedure is explained by Friedman. [FR1] [SINK]

"ACA and You" presents general properties of English letters. [ACA]

Foster presents detail letter characteristics based on the Brown Corpus. [CCF]

Don L. Dow puts out a clever computer cryptogram game which does frequency analysis and is user friendly for very simple Aristocrats. {Available as shareware} [DOW]

Depending the basis text we choose, we find variations in the frequency of letters. For example, literary English gives slightly different results than frequencies based on military or ordinary English text.

Hagn presented Literary English Letter Usage Statistics based on "A Tale of Two Cities" by Charles Dickens as follows: [HAGN]

```
Total letter count =  586747
Letter use frequencies:     Total doubled letter count = 14421
E:    72881    12.4%        Doubled letter frequencies:
T:    52397     8.9%        LL:     2979    20.6%
A:    47072     8.0%        EE:     2146    14.8%
O:    45116     7.6%        SS:     2128    14.7%
N:    41316     7.0%        OO:     2064    14.3%
I:    39710     6.7%        TT:     1169     8.1%
H:    38334     6.5%        RR:     1068     7.4%
S:    36770     6.2%        PP:      628     4.3%
R:    35946     6.1%        FF:      430     2.9%
D:    27487     4.6%        NN:      301     2.0%
L:    21479     3.6%        CC:      243     1.6%
U:    16218     2.7%        MM:      207     1.4%
M:    14928     2.5%        DD:      201     1.3%
W:    13835     2.3%        GG:       99     0.6%
C:    13223     2.2%        BB:       41     0.2%
F:    13152     2.2%        ZZ:       13     0.0%
G:    12121     2.0%        AA:        2     0.0%
Y:    11849     2.0%        HH:        1     0.0%
P:     9452     1.6%
B:     8163     1.3%
V:     5044     0.8%
K:     4631     0.7%
Q:      655     0.1%
X:      637     0.1%
J:      623     0.1%
Z:      213     0.0%
```

```
Total initial letters =  135664  Total ending letters =  135759
Initial letter frequencies:      Ending letter frequencies:

T:    20665    15.2%             E:    26439    19.4%
A:    15564    11.4%             D:    17313    12.7%
H:    11623     8.5%             S:    14737    10.8%
W:     9597     7.0%             T:    13685    10.0%
I:     9468     6.9%             N:    10525     7.7%
S:     9376     6.9%             R:     9491     6.9%
O:     8205     6.0%             Y:     7915     5.8%
M:     6293     4.6%             O:     6226     4.5%
B:     5831     4.2%             F:     5133     3.7%
C:     4962     3.6%             G:     4463     3.2%
F:     4843     3.5%             H:     3579     2.6%

Top digraphs:
TH:   17783    RE:   8139   ED:   6217   IS:   5566
HE:   17226    ND:   7793   AT:   6200   NG:   5564
IN:   10783    HA:   6611   EN:   5849   IT:   5559
ER:   10172    ON:   6464   HI:   5730   OR:   4915
AN:   9974     OU:   6418   TO:   5703   AS:   4836
```

POSITION AND FREQUENCY TABLE

Time to put to good use the barrage of data presented. Given the next slightly harder cryptogram, and ignoring again a pattern word attack, we can develop some useful tools. [Much of what I am covering can be done automatically by computer but then your brain goes mushy for failure to understand the process.]

A-2. [no clue] S-TUCK

V W H A Z S J X I H S K I M F M W C G M V W O J S I F A G F J A Q

Q M N R J K Z M G R S W M F. J A T W X H A W F. F I Q Q W F F X I H

F K H B A O Z J S M A H H F. T G A H P K D X M A W O V F S A R F

X H K I M A F S.

First we perform a CT Frequency Count.

```
 F  A  H  M  W  S  I  J  K  X  G  Q  O  R  V  Z  T B C D N P
13 11  9  9  8  7  6  6  5  5  4  4  3  3  3  3  2 1 1 1 1 1
```

We have 106 letters. 20% are considered low frequency. 20% of 106 = 21. Counting from right to left we have O, R, V, Z, T, B, C, D, N, P. We mark A-2. with a dot over each appearance. We also enter the frequency data under the CT.

Vowels contact the low frequency letters more often than do consonants. About 80% of the time. We use S-TUCK method combined with our text. [ELCY] [TUCK]

We go thru A-2. writing down the contact letters on both sides, for low frequency CT. We tally one for each contact. If a CT letter is between two low frequency letters we tally 2. Contacts for low frequency letters touching each other = 0. We do not count N o R in word 2, and in word 1, W contacts V, so W is tallied with 1. A an S contact Z, so both A and S are credited. We get:

```
/////  ////  //   ///   ///   //   ///   //   //
 W      A     S    G     M    J     K    H    F
```

Low Frequency Contacts for A-2.

From the Brown Corpus, vowel contact as percentage of total number of digrams is low: [CCF]

```
                Second
            A   E   I   O   U   Y

       A    0   0  .4   0   .1  .3
                                      Total nonpairs = 5.1%
       E   .7  .4  .2  .1   0   .2              pairs = 0.7%
  F
  I    I   .2  .4   0  .7   0    0
  R
  S    O   .1  .1  .1  .3   1.0  0
  T
       U   .1  .1  .1  0    0    0

       Y  0   .1   0  .2   0    0
```

ELCY tells us quite a bit about vowel behavior.

1. A, E, I, O, are normally high frequency, U is moderate and Y is low frequency.

2. Letters contacting low frequency letters are usually vowels.

3. Letters showing a wide variety of contact-letters are usually vowels.

4. In repeated digrams, one letter is usually a vowel.

5. In reversed digrams, one letter is usually a vowel.

6. Doubled consonants ar usually flanked by vowels, and visa versa. (cvvc or vccv)

7. It is unusual to find more than 5 consonants in succession.

8. Vowels do not often contact each other.

9. If the CT letter with highest frequency is assumed E, any other high frequency letter which never touches E, can be assumed a vowel. A letter that contacts it very often can not be a vowel.

10. E is most frequent vowel and rarely touches O. Both double freely.

11. The vowel that follows and rarely precedes E is A.

12. The vowel that reverse with E is I.

13. Observations 11 and 12 apply to the vowel O. However, finding U it precedes E and follows O.

14. The only vowel-vowel digrams of consequence are OU,EA,IO.

15. Three vowels in sequence may be IOU, EOU, UOU, EAU.

NYPHO's Robot says that the first four or last four letters of a word contain a vowel. [TUCK]

ELCY defines high frequency letter behavior.

About 70% of the language is made up of E, T, A, O, N, I, R, S, H. This high frequency group has three cliques.

Class I. T, O, S appear frequently both as Initials and Finals; terminal O in short words like to. All double freely

Class II. A, I, H appear frequently as initials, but rare as finals, especially A, I. They do not readily double.

Class III. E, N, R, appear frequently as finals, less frequently as initials, frequently double, especially E, N and R not so often.

When one of these letters changes its class, the least likely exchange is one occurring between Class II and III.

ELCY gives us tips for identifying consonants:

1. Those letters still remaining in the high frequency section will usually include T, N, R, S, H. H is the easiest to identify, it precedes all vowels, and forms TH, HE, HA.

2. R is also recognizable with it reverses openly with all vowels, and links with the class I club.

3. T is usually found by frequency, precedes vowels rather than follow them, precedes consonants. S has a similar pattern to a lesser degree. N confuses this picture.

4. ST -TS AND RT -TR are the only frequent consonant reversals.

5. TT and SS are most frequent doubles in language.

Having all this information, we are well armed against even the most resistant Aristocrat.

We return now to solution of A-2.

From the number of their contacts, W and A are most likely vowels. G, K, M are next most likely.

We look at these letters in the position table.

W. has the looks of E even though it is not the most frequent.

A. cannot be A so it might be I. but frequency may be too high.

G. and K. have inside positions and look like vowels but can not be identified.

M. might be O by frequency but is confused with R.

A study of A-2. shows that W and A reverse which might be ei and ie. AG reverses which might be io or ia. M repeats, and reverses with W and G. It most likely is R not O. K does not contact W A G or M. We mark the cipher with W A G K as vowels and M as a consonant, putting in the assumed values.

A-2. [no clue] S-TUCK

```
        1               2           3                   4
d e l i g h t f u l   h o u r s   r e   a r d   e   t h   s i a s t i c
. v c v .     c v c     v v c c   c v . v c .   v .     v c v v c   v c
V W H A Z S J X I H   S K I M F   M W C G M V   W O J S I F A G F J A Q
3 8 9 + 3 7 6 5 6 9   7 5 6 9 *   9 8 1 4 9 3   8 3 6 7 6 * + 4 * 6 + 4

          5                   6         7                   8
c r     t o g r     h e r s     t i       f l i e s     s u c c e s s f u l
c c . .   v . c v .   v c c       v . v   c c v v c     c v c c v c c c v c
Q M N R J K Z M G R S W M F.   J A T W   X H A W F.     F I Q Q W F F X I H
4 9 1 3 6 5 3 9 4 3 7 8 9 *   6 + 2 8   5 9 + 8 *     * 6 4 4 8 * * 5 6 9

      9           10              11                12
s o l   i   g   t h r i l l s     a i l   o     f r i e   d s h i   s
c v c . v . .       c v c c c   . v v c . v .   c c v v . . c   v . c
F K H B A O Z   J S M A H H F.  T G A H P K D   X M A W O V F S A R F
* 5 9 1 + 3 3   6 7 9 + 9 9 *   2 4 + 9 1 5 1   5 9 + 8 3 3 * 7 + 3 *

      13
f l   u   i s h
c c v v c v c
X H K I M A F S.
5 9 5 6 9 + * 7                          (two digit figures F=13=* ; A=11=+)
```

Using Nympho' robots rule, in Word 1, J X I H, one must be a vowel. Word 8 shows F X I H contains a vowel. Word one suggest the ending 'ful'. X = f and H = l. Examine X I H and the I is in the vowel positions. (inner positions). So the vowels are now W E G K I. From its end position F =s. In words 4 and 11, GA reverses so G cannot be a u for ui is not a reversal. We try KI=ou, therefore G = A. Put into the above cipher tableaus. Word 5 breaks the two c's, so Q = c.Word 1 might be delightful, so V=d, ZSJ = ght. Remember the second letter position favors vowels. [ROBO]

The message reads: Delightful hours reward enthusiastic cryptographers. Time flies. Successful solving thrills. Mailbox friendships flourish. KW =K1=salutory.

PATTERN WORD ATTACK

Pattern words are words for which one or more letters are repeated such as awkward, successful, interesting, unusually. Aegean Park Press publishes pattern word books from 3 - 16 letters. Pattern words lists are indexed by key letters or figures or by vowel consonant relationships. [BARK] Pattern words give a quick wedge into the cryptogram. One of the best Pattern Word Dictionaries is the Cryptodyct. [GODD]

The Crypto Drop Box has the TEA computer program which gives automated pattern searching and anagraming up to 20 words. It is a very effective tool.

In A-2. We find a prize in word 8. Using a key letter approach:

```
A B C C D A A E B F
F I Q Q W F F X I H
```

or

```
1 2 3 3 4 1 1 5 2 6   = (334) 11526 [10L]
F I Q Q W F F X I H
```

The first pattern found on page 310 Appendix of [CCF] is successful. The Cryptodyct uses the latter indexing method and under 10 letter words we find that the 334 11526 pattern equals successful.

Cryptographers generate their own special lists:

Transposals: from, form; night, thing; mate, meat;
Queer words: adieu, crwth, eggglass, giaour, meaow
Consonant sequences: dths, lcht, ncht, rids, ngst, rths
Favorite ins: people, crypt, success,

Using the TEA model, it was necessary to assume the vowels at u and e for a 1u22e445u6 template to get successful and juggernaut on the first try.

Non Pattern word lists are those with words that do not have even one repeated letter, such as come, wrath, journey. They are very useful in attacking Patristrocrats and very difficult Risties.

OMAR gave us this fine list in order of frequency:

CRYPT	WORDS	ABOUT	KNOWS	BELOW	OKAPI	SWORD
BLACK	ALONG	AFTER	NEGRO	EXTRA	PLACE	THREW
WATCH	CRAZY	CAUSE	UNDER	FIRST	SIXTY	WRONG
WHILE	CROWD	DRUNK	UPSET	FOUND	STUDY	
ANGRY	PLUMB	EMPTY	YIELD			

We will come back to it in the Patty section.

Also in the CDB is a program called ASOLVER which automates the Digram solution method to get the best fit.

MORE ABOUT VOWEL POSITION PREFERENCES

Dr. Raj Wal summarized Barkers Vowel Preferences data. He also developed cross correlation coefficients for each letter. Foster details this work in his book. [CCF]

This handy little table gives us an entry when needed. It is correct more times than it fails.

Word Length	Position Preferences							
one	1							
	V							
two	1	2						
	V	C						
three	1	2	3					
	C	C	-					
four	1	2	3	4				
	C	V	-	C				
five	1	2	3	4	5			
	C	C	V	C	C			
six	1	2	3	4	5	6		
	C	V	C	-	-	C		
seven	1	2	3	4	5	6	7	
	C	V	C	C	-	-	C	
eight	1	2	3	4	5	.	.	Final
plus	C	C	-	-	-	-	-	C

Variety of Contact Table (VOC):

```
Freq: 8  7  6 5 4 4 6 5 4 7  /  3 3 6 3  /  2 1 1 1 1 1
VOC:  10 9  8 8 7 7 7 6 6 6  /  5 5 5 5  /  4 2 1 1 1 1
CT:   X  Y  B V W K A U H Z  /  M R C S  /  T N E O D P
```

We start with the position that 20% of the text represented by variety count are consonants. 20% of 104 = about 21. The line of demarcation is between R and C but 4 letters have the same VOC of 5, M,R,S,C. If we take one , we must take all and one of these most likely is a vowel. The key to solution is the VOC "step up" versus "step down" observation. Vowels tend to step up and Consonants tend to step down. [i.e. 3M5 is a step up of 2 points and 6C5 is a step down of one point.]

M, R, S all step up, C steps down 1 point and most likely is a consonant. We develop a separation line and place the contacts on each side of the consonant line starting from the right of the VOC table.

```
First Consonant Line
   C T N E O D P
--------------------
      V |
      X | XXXX
     YY | YYY
      K |
      S |
      Z |
        | W
        | R
      M |
        | H
        | B
```

If any letter does not appear at all below the line, that letter is most likely a consonant. A and U fall into this catagory. We add these to analysis:

```
Second Consonant Line
  C T N E O D P A U
--------------------
       VV | V        mark X and Y as Vowels
        X | XXXX     (vowel)  both step up
     YYYY | YYY      (vowel)  with high VOC
      KKK |
        S |
        Z | ZZ      consonant (step down)
          | WWW     test as h
        R | R
       MM |
          | HHH
        B | B
          | U
        A |
          |
```

We shift to A-3 and mark in the suspected consonents.

A-3. No clue. Author Bosley No. 19. CM. June 1936.
cont

```
       1               2                 3                  4

U W Y M N X K A    E H X R B Z      U V X M U W B Z   O Y Z T W H V C X Y A
- - o - - o --     - o o - o -      - - o - - - o -   - o - - - o - - o o -

    5               6                    7                  8

C Y A U Z    D B R A H V K B A;   Z W S V A H K U Z B K C,     M S C X
- o - - -    - o - - o - - o -    - - o - - o - - - o - -      - o - o

    9              10

C Y X B S,   X V Z Y T R Y C X P.                                        (104L)
- o o o o    o - - o - - o - o -
```

n and h turn up on the right and left side of the consonant line freely. w and h are candidates. Since h=H, then w might equal h. Digrams such as sh or ch are prevalent. W is the second position in word 7 which tentatively confirms the PT h and suggests that Z is a consonant (step down). B is astep up as well as S. The third word confirms but the 9 word has four vowels. Hmm? K and H are both possibilities for vowels. Word 4 tends to favor the H. So:

```
Final Consonant Line
C T N E O D P A U W Z
---------------------
     VVV | V         mark X and Y as Vowels
       X | XXXX      (vowel)  both step up
   YYYYY | YYYYY     (vowel)  with high VOC
     KKK |
       S | S        vowel low freq? =u?
      ZZ | ZZ       consonant (step down)
         | WWWW     test as h
       R | R
      MM |
         | HHHH
     BBB | BBB       vowel
    UUUU | U         consonant
       A |           consonant
       T | T         consonant
```

Let me fill in where ELCY stops. A-3 has vowels and consonants separated. We have the PT letter h. Word 9 is either clever or wrong. Using Barkers Pattern List on p39, we find bayou and miaou. The same reference gives us thunderclaps for word 7. Although not correct we find thunderstorm matching the pattern under 819710/12W and word 8 suggests puma. The final message reads: shipyard zealot snapshot kitchenmaid midst goldenrod; thunderstorm, puma miaou, anticlimax.

The TEA database yields words: thunderstorm and anticlimax. The reader is invited to reconstruct the keywords, if any.

NON-PATTERN WORD ATTACK

Try this Aristocrat.

A-4. Fire, fire burning bright. by Ah Tin Dhu.

```
   1          2          3          4          5          6          7
A B C D E  A C F G H  I C J F H  K C I B L  K F B H L  K C M J N  O M J P I

   8          9          10         11         12         13         14
B H L M C  M R S P E  B C A I H  T I A U H. K U M C E  V D U H P. S C F G D

   15         16         17         18         19
J W B I L  J S U M L  D U V N P, V E O M L  C F G L E.
```

To solve by using non-pattern words, 3 or 4 words in the cipher having several letters in common. Under one of these write 5 or 6 words from the pattern list. We will use OMAR's list given previously. Note the initials and final letters and letter positions of the trial words. In A-4. K is an initial and L is a terminal. Choose the non-pattern words to conform with this requirement. We write the common letters under the trial word and try to make clear message out of the balance of CT. Word 5 has K, BHL and F.

```
  K F B H L    A C F G H    K C I B L    B H L M C
1 b l a c k      l  c    b    a k    a c k
2 c r a z y      r  z    c    a y    a z y
3 w r o n g      r  n    w    o g    o n g
4 c r o w d      r  w    c    o d    o w d
5 d r u n k      r  n    d    u k    u n k
6 f o u n d      o  n    f    u d    u n d
```

Line 6 arson, fraud, under. Putting this into the risties we get:

```
   1          2          3          4          5          6          7
b u r  y   b r o w n  a r s o n  f r a u d  f o u n d  f r e      e    a
A B C D E  A C F G H  I C J F H  K C I B L  K F B H L  K C M J N  O M J P I

   8          9          10         11         12         13         14
u n d e r  e     y  u r b a n  c a b i n  f i e r y    i n      r o w
B H L M C  M R S P E  B C A I H  T I A U H. K U M C E  V D U H P. S C F G D

   15         16         17         18         19
   u a d      i e d    i        y   e d  r o w d y
J W B I L  J S U M L  D U V N P, V E O M L  C F G L E.
```

All the vowels are id'ed and r, n. The message is "Burly brown arson fraud found fresh vesta under empty cabin. Fiery glint. Prowl squad spied light, gyved rowdy."

RECAP

1. Common letters appear frequently in a message but not necessarily in exact correspondence to the uniform frequency distribution.

2. Start working with shorter words, common endings.

3. Look for repetitions of bigrams, trigrams, reversals.

4. Go with the flow of the cipher text and extract all the information on frequency, position and contacts.

5. Eliminate all but few possibilities. Test and confirm. Test and Confirm.

6. Work back and forth from the cryptogram and the keyword alphabets. Expect the message to make some kind of sense.

7. Look for patterns or non patterns. Separate vowels and consonants. Try brute force. Use lists.

8. Persevere.

CM REFERENCES

PHOENIX has compiled a list of articles (page 2) concerning ARISTOCRATS between 1932 - 1993 in "The Cryptogram Index," available through the ACA. On page 27, he lists additional references on simple substitution. Articles by B.NATURAL and S-TUCK are especially useful. [INDE]

HOMEWORK PROBLEMS

Solve these cryptograms, recovery the keywords, and send your solutions to me for credit. Be sure to show how you cracked them. If you used a computer program, please provide "gut" details. Answers do not need to be typed but should be generously spaced and not in RED color. Let me know what part of the problem was the "ah ha", i e. the light of inspiration that brought for the message to you.

A-1. Bad design. K2 (91) AURION

V G S E U L Z K W U F G Z G O N G M V D G X Z A J U X U V B Z

H B U K N D W V O N D K X D K U H H G D F N Z X U K Y D K V G U N

A J U X O U B B S X D K K G B P Z K D F N Y Z B U L Z .

A-2. Not now. K1 (92) BRASSPOUNDER

K D C Y L Q Z K T L J Q X C Y M D B C Y J Q L : " T R H Y D F K X C ,

F Q M K X R L Q Q I Q H Y D L M K L D X C T W R D C D L Q

J Q M N K X T M B P T B M Y E Q L K F K H C Y L Q Z K T L T C . "

A-3. Ms. Packman really works! K4 (101) APEX DX

* Z D D Y Y D Q T Q M A R P A C , * Q A K C M K * T D V S V K . B P W V G

Q N V O M C M V B : L D X V K Q A M S P D L V Q U , L D B Z I U V K Q F

P O W A M U X V , E M U V P X Q N V , U A M O Z N Q K L M O V

(S A P Z V O) .

A-4. Money value. K4 (80) PETROUSHKA

D V T U W E F S Y Z C V S H W B D X P U Y T C Q P V E V Z F D A E S T U W X

Q V S P F D B Y P Q Y V D A F S , H Y B P Q P F Y V C D Q S F I T X P X B J

D H W Y Z .

A-5. Zoology lesson. K4 (78) MICROPOD

A S P D G U L W , J Y C R S K U Q N B H Y Q I X S P I N

O C B Z A Y W N = O G S J Q O S R Y U W , J N Y X U O B Z A (B C W S

D U R B C) T B G A W U Q E S L. * C B S W

REFERENCES

[ACA] ACA and You, Handbook For Members of the American Cryptogram Association, 1995.

[BARK] Barker, Wayne G., "Cryptanalysis of The Simple Substitution Cipher with Word Divisions," Aegean Park Press, Laguna Hills, CA. 1973.

[BAR1] Barker, Wayne G., "Course No 201, Cryptanalysis of The Simple Substitution Cipher with Word Divisions," Aegean Park Press, Laguna Hills, CA. 1975.

[B201] Barker, Wayne G., "Cryptanalysis of The Simple Substitution Cipher with Word Divisions," Course #201, Aegean Park Press, Laguna Hills, CA. 1982.

[BP82] Beker, H., and Piper, F., " Cipher Systems, The Protection of Communications", John Wiley and Sons, NY, 1982.

[CCF] Foster, C. C., "Cryptanalysis for Microcomputers", Hayden Books, Rochelle Park, NK, 1990.

[DOW] Dow, Don. L., "Crypto-Mania, Version 3.0", Box 1111, Nashua, NH. 03061-1111, (603) 880-6472, Cost $15 for registered version and available as shareware under CRYPTM.zip on CIS or zipnet.

[ELCY] Gaines, Helen Fouche, Cryptanalysis, Dover, New York, 1956.

[GODD] Goddard, Eldridge and Thelma, "Cryptodyct," Marion, Iowa, 1976

[FR1] Friedman, William F. and Callimahos, Lambros D., Military Cryptanalytics Part I - Volume 1, Aegean Park Press, Laguna Hills, CA, 1985.

[FR2] Friedman, William F. and Callimahos, Lambros D., Military Cryptanalytics Part I - Volume 2, Aegean Park Press, Laguna Hills, CA, 1985.

[FRE] Friedman, William F. , "Elements of Cryptanalysis," Aegean Park Press, Laguna Hills, CA, 1976.

[HA] Hahn, Karl, " Frequency of Letters", English Letter Usage Statistics using as a sample, "A Tale of Two Cities" by Charles Dickens, Usenet SCI.Crypt, 4 Aug 1994.

[INDE] PHOENIX, Index to the Cryptogram: 1932-1993, ACA, 1994.

[NIC1] Nichols, Randall K., "Xeno Data on 10 Different Languages," ACA-L, August 18, 1995.

[NIC2] Nichols, Randall K., "Chinese Cryptography Part 1," ACA-L, August 24, 1995.

[OP20] "Course in Cryptanalysis," OP-20-G', Navy Department, Office of Chief of Naval Operations, Washington, 1941.

[ROBO] NYPHO, The Cryptogram, Dec 1940, Feb, 1941.

[SINK] Sinkov, Abraham, "Elementary Cryptanalysis", The Mathematical Assoc of America, NYU, 1966.

[STIN] Stinson, D. R., "Cryptography, Theory and Practice," CRC Press, London, 1995.

[TUCK] Harris, Frances A., "Solving Simple Substitution Ciphers," ACA, 1959.

Notes

Throughout my lectures, PT will be shown in lower case. CT will be shown in upper case. As a convention, Plain text will generally be shown above the Cipher text equivalent.

A = Aristocrats, P = Patristrocrats, X = Xenocrypts

Any typo errors are my responsibility. I probably fell asleep at the keyboard. Please advise and I will correct them as well as put out an erratum sheet at the end of the course. Students may want to start a 3" permanent binder with separators for the various lectures and materials.

OUTLINE

1. Intro - First Principles - Global Mathematical Nature
2. Keyword Systems and Conventions Used
3. Simple Substitution Cryptanalysis without/with Complexities

    a. Eyeball
    b. Frequency Distributions - General Nature of English Letters
    c. Friedman Techniques - Random vs Expected -Spaces and a Wealth of Tables: Digram, Trigram, and more
    d. C. C. Foster Techniques
    e. S-Tuck Techniques
    f. Pattern Words
    g. ELCY : Consonant Line Attack
    h. Sinkov Techniques
    i. Barker's Vowel Separation and Position Table
    j. Non Pattern Words: "Dooseys"
    k. SI SI Patterns
    l. CM References for Risties
    m. Relationship to XENOS:French and German Solutions
    n. Computer Program Aids - TEA Database, CDB, ABACUS, Computer Supplement
    o. References

4. Homework Problems
5. Variant Substitution Systems

    a. Friedman
    b. Waxton

Next lecture we will cover the balance of the outline material and jump into Patristocrats.