Spring 2015
Chris Christensen
CSS/MAT 483

Columnar Transposition

Classically ciphers that rearranged the letters of plaintext were called transposition ciphers. They can be recognized because ciphertext letter frequencies are the same as plaintext letter frequencies.

Columnar transposition is probably the most commonly studied transposition cipher. We will use that method to encrypt the following "pilot's saying:"

`The nose is pointing down and the houses are getting bigger.`

There are 49 letters in the message. We want to place the letters of the message in a rectangular array. In this case, because we would like the rectangular array to have 49 cells, a 7×7 array may be used. We also need a keyword having its length the same as the number of columns – we will use *analyst*.

A	N	A	L	Y	S	T
1	4	2	3	7	5	6
t	h	e	n	o	s	e
i	s	p	o	i	n	t
i	n	g	d	o	w	n
a	n	d	t	h	e	h
o	u	s	e	s	a	r
e	g	e	t	t	i	n
g	b	i	g	g	e	r

The ciphertext is obtained by reading down the columns in the order of the numbered columns (which are alphabetically ordered).

`TIIAOEGEPGDSEINODTETGHSNNUGBSNWEAIEETNHRNROIOHSTG`

Our message exactly fit the rectangular array. If the message does not completely fill the array, nulls (i.e., padding) may be added to fill the array (this is the easier cipher to break) or not (this is harder to break because the columns do not all have the same length). In the latter case, the length of the keyword determines the number of columns, and the number of letters in the message determines the number of complete and partial rows.

The transposition should be applied several times if the plaintext message were longer than 49 letters.

Remember, for encrypting, "in by rows and out by columns."

Decrypting the columnar transposition

Here is a message that was encrypted using a rectangular array with keyword *analyst*.

```
TRLEELIGCIGEHALANTNCTECYENEN
```

Because the keyword has 7 letters, we know that the rectangular array has 7 columns. The message has 28 letters; therefore, the array must be 4×7. Each column must have 4 entries.

First, we place the letters of the keyword in alphabetical order: *aalnsty*. Then place the ciphertext letters in columns.

A	A	L	N	S	T	Y
t	e	c	h	n	t	e
r	l	i	a	t	e	n
l	i	g	l	n	c	e
e	g	e	a	c	y	n

Now rearrange the letters of the keyword to form *analyst*.

A	N	A	L	Y	S	T
t	h	e	c	e	n	t
r	a	l	i	n	t	e
l	l	i	g	e	n	c
e	a	g	e	n	c	y

The plaintext message is `the central intelligence agency`. (Notice that there could be some ambiguity about which "A" column comes first. We have used the convention that the first "A" column will correspond to the first *a* in *analyst*.)

Remember, for decrypting, “in by columns and out by rows.”

Cryptanalysis of the columnar transposition

We will do only "the easy case;" i.e., we will assume that the columnar transposition uses a rectangular array that was completely filled.

Here is the ciphertext:

```
ASAIR  ITFNM  IMTKL  SOIEE  M
```

The “key” to cryptanalyzing the ciphertext is to determine the number of columns; i.e., the length of the keyword. There are 21 letters in the ciphertext. Because we know that the message completely fills the rectangle, this suggests either a 3×7 or a 7×3 array.

We arrange the ciphertext in columns.

Either

A	I	T	M	T	S	E	
S	R	F	I	K	O	E	
A	I	N	M	L	I	M	

or

A	F	L
S	N	S
A	M	O
I	I	I
R	M	E
I	T	E
T	K	M

.

The solution is by anagramming (making a word or portion(s) of word(s) by rearranging letters) a row.

The 7×3 arrangement seems unlikely because it has a string `TKM` with no vowels that is unlikely. Also, the `III` is unlikely. So, let us try the 3×7 arrangement. Notice that there are 7! = 5040 arrangements of the columns. We would like to not have to try all of them!

A	I	T	M	T	S	E
S	R	F	I	K	O	E
A	I	N	M	L	I	M

In the first row, `MATE` seems to leap out. This leaves `ITS`. Perhaps, a slightly wrong guess – `ESTIMAT-` seems to be a possibility.

Let us rearrange the columns.

E	S	T	I	M	A	T
E	O	K	R	I	S	F
M	I	L	I	M	A	N

Not quite, but there are two Ts in the first row. Let us swap those columns.

E S T I M A T
E O F R I S K
M I N I M A L

This works. Notice that because we have multiple rows that are permuted the same way, we can use multiple anagramming for cryptanalysis.

It is often worthwhile to write the ciphertext in columns, cut out the columns, and rearrange the columns to do the anagramming.

Determining the dimension of the rectangle

Frequencies can help to determine the dimensions of the rectangle. In English approximately 40% of plaintext consists of vowels. Therefore, for the correct dimension, each row of the rectangle should be approximately 40% vowels. Consider our choice between 3×7 and 7×3.

For a 3×7 rectangle, each row should contain approximately 2.8 vowels. Let us note the difference between this estimate and the actual count:

							Number of vowels	Difference
A	I	T	M	T	S	E	3	0.2
S	R	F	I	K	O	E	3	0.2
A	I	N	M	L	I	M	3	0.2

The sum of the differences is 0.6.

For a 7×3 rectangle:

			Number of vowels	Difference
A	F	L	1	0.2
S	N	S	0	1.2
A	M	O	2	0.8
I	I	I	3	1.8
R	M	E	1	0.2
I	T	E	2	0.8
T	K	M	0	1.2

The sum of the differences is 6.2. It appears that the 3×7 rectangle is more likely.

Using digraph frequencies to arrange the columns

Digraph frequencies can be used to help in the cryptanalysis in place of just looking for reasonable pairings of the columns. For example, consider our ciphertext above `ASAIR ITFNM IMTKL SOIEE M`. Again, we'll assume that a 3×7 rectangle is appropriate.

A	I	T	M	T	S	E
S	R	F	I	K	O	E
A	I	N	M	L	I	M

We will pair the first column with each of the other columns on the right and consider how likely it is that such digraphs will occur in English. The frequencies we will use come from Sinkov (see the appendix). Recall that there are $26 \times 26 = 676$ digraph frequencies.

AI	311	AT	1019	AM	182	AT	1019	AS	648	AE	13
SR	9	SF	8	SI	390	SK	30	SO	234	SE	595
AI	311	AN	1216	AM	182	AL	681	AI	311	AM	182
	631		2243		754		1730		1193		790

The most likely pairing is

AT
SF
AN

Oops! We know that this is not the correct pairing, but the second most likely pairing is correct. (During cryptanalysis, we don't always get the correct result on the first try.)

Once we have a pairing, we could then continue using digraph frequencies to select columns to add on the left and on the right. Etc.

More columnar transposition

It would be harder to do the cryptanalysis if the rectangle were not completely filled. For example, let's use a columnar transposition with keyword *norse* to encrypt the message `Germany seeks an alliance`. The message contains 22 letters; so, we need 4 complete rows and one partial row.

In by rows:

n	o	r	s	e
g	e	r	m	a
n	y	s	e	e
k	s	a	n	a
l	l	i	a	n
c	e			

Out by columns:

`AEANG  NKLCE  YSLER  SAIME  NA`

Because the columns do not have the same length, this would not be as easy to cryptanalyze. It would not be obvious how many columns were used. (The size of the rectangle would be either 2×11 or 11×2 if we knew that a full rectangle had been used; i.e., the keyword would have length either 11 or 2.)

However, if we know the keyword, decrypting is no problem. Try decrypting the ciphertext

`IMYRA  CBILM  AANIE  NSBNR  ESE`

which was encrypted with columnar transposition with keyword *norse*.