2.4 Lesson 3 : Substitution ciphers

Ciphers can be divided into two branches, known as transposition and substitutions.
In transposition, the letters of the message are simply rearranged, generating an anagram or a table. One decodes the ciphertext by rearranging back the letters.
In a substitution cipher each letter (or symbol) is represented by other letters. The ciphertext can be deciphered by anyone knowing the order of the cipher alphabet used

Exercise 13 — Caesar's shift, . See answers

Julius Caesar, Roman general and statesman, invented a cipher to encode messages send to his generals. In a "Caesar's shift of 3" **each letter is replaced with the letter that is 3 places further down the alphabet**

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

1. Encode the message "Julius Caesar was not emporor" using a Caesar shift of 3.

   .....

2. What does the following message say? "EDFN DW VHYHQ" .....
3. How can we design a different Caesar cipher ? .....

   How many ways are there of doing this ? .....

4. The Vigenere square below shows all possible amounts of possible shifts.

   The following message uses one of the shifted alphabets from the Vigenere square. What does it say?

   BPQA PIA JMMV APQNBML JG MQOPB

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
0	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
1	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	A
2	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B
3	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C
4	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D
5	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E
6	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F
7	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G
8	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H
9	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I
10	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J
11	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K
12	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L
13	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M
14	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N
15	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
16	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
17	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
18	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
19	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
20	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
21	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
22	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
23	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
24	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
25	Z	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

Going through all possibilities one by one to see which makes sense is a tedious way of doing it! We can use the fact that in English, some letters occur more often than others. In the English language, the most common letter is "E"

Exercise 14 — Breaking the code. See answers

Using a Caesar's shift cipher produced the following message :

VXKT BT RWTTHT EATAHT

1. Use tally table to determine the frequency of each letter.

Ciphertext	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
frequency																										

2. Which letter occurs most often in the coded message ?

3. What letter might this represent? .....How much of a shift is this ?

4. Can you use this fact to decode the message without trying every possibility.

Ciphertext	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
Plaintext																										

Exercise 15 — General case. See answers

In a Caesar's shift, the coded alphabet is in order. If the coded alphabet is not in order, then we have a **substitution cipher**. For example :

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

What does this message say?

ZHCVHYP NYDCTG SJL GCMO

It may seem that arbitrary substitution ciphers would be very hard to undo without knowing the order of the cipher alphabet. However, we can use the same trick used in exercise **??**. *This process we now call* ***Frequency Analysis*** *was first described by Arab polymath Al-Kindi. His "A Manuscript on Deciphering Cryptographic Messages" (lost until found in 1987) was quoted for over seven centuries.*

Exercise 16 — 👂. See answers

Watch "The Science of Secrecy" by Simon Lehna Singh (5'00")

1. When was **Frequency Analysis** discovered by Al-Kindi?

2. What letters are the first to be guessed in a ciphertext?

3. What are the requirements for a frequency analysis, and when will it fail to undo the cipher? Explain.

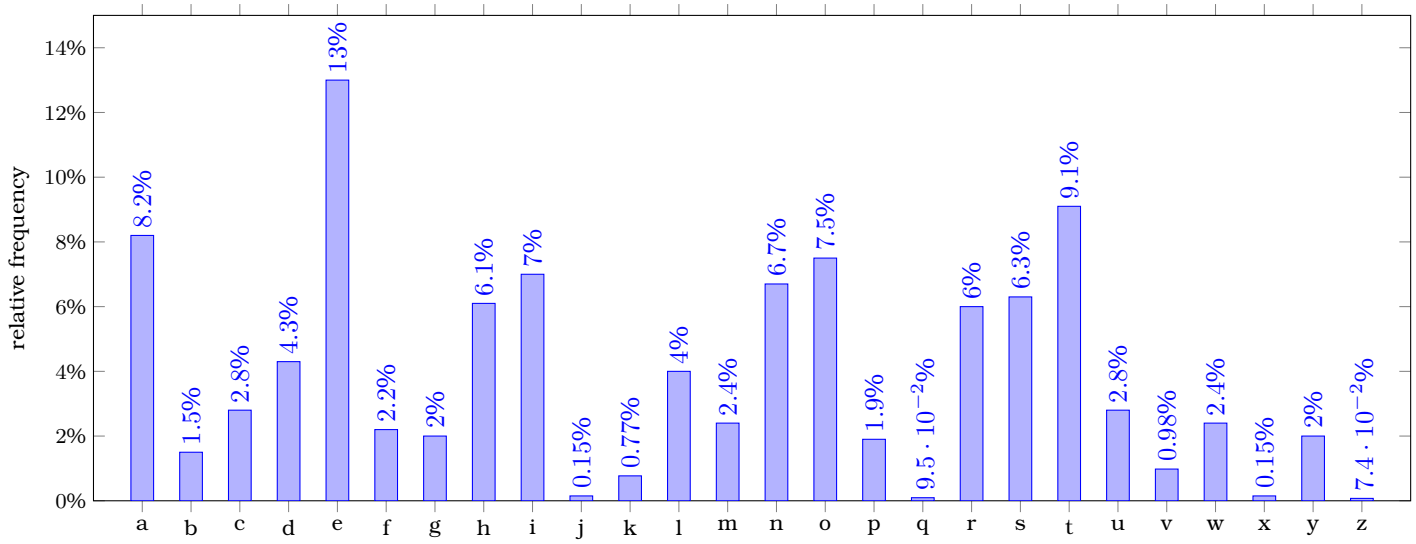

Figure 2.1: The most common letters in text are, in order, ETAON RISHD LFCMU GYPWB VKJXZQ.

Exercise 17 — Breaking the code! (part 2). See answers

The following text uses yet another substitution cipher. Your task is to decipher the passage and complete the substitution alphabet. Several hints are given to help you.

AUHC MVKFC V BYZUGC V IZMC CJ GUMBZYAZD UKUVM VC HZZGZB CJ GZ V HCJJB PD CFZ VYJM KUCZ AZUBVMK CJ CFZ BYVWZ UMB OJY U IFVAZ V TJNAB MJC ZMCZY OJY CFZ IUD IUH PUYYZB CJ GZ.

Ciphertext	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
tally																										
frequency																										
Plaintext																										

Hint №1: The three most frequently occurring letters in the passage concur with the three most frequent letters in the English language. Find the three most commonly occurring letters in the cipher and substitute the letters you think they could represent.

Hint №2: Note that there are some one-letter words; one of these you should already have found. What would the other one be? Use this information to find a fourth letter.

Hint №3: The next most frequently occurring letter in the cipher can now be assigned its real letter, So you now have a fifth letter

Hint №4: If you have done everything correctly, you should have a couple of words that look like “T?E”. Use this information to find a sixth letter.

Hint №5: Look at the word “?ATE”. There are a few possibilities: DATE, FATE, GATE, LATE, MATE, RATE, SATE. Note that whatever the letter K stands for it stands for the same thing in the second word “?I?HT”.

Hint №6: Word 20 has a very common ending. By now you should have enough to work out/guess (a very important skills in cipher analysis) to decipher the whole message!

Hint №7: Once you have deciphered the whole message, are you able to give the complete substitution table? If not, what would you need to finish the task?