# TTM4135 Practical

## 1 Background

This assignment consists of practical experiments with historical ciphers. The task is to analyse four different ciphertexts, identify the cipher used, and find (partially) the plaintexts.

The original plaintexts are all written in English from one or more novels written 70-80 years ago. Each plaintext is different and each is encrypted with a different classical cipher algorithm. The four algorithms used, in random order in your set, are:

- random simple substitution
- Caesar
- Vigenère
- 2 x 2 Hill cipher

For all of the ciphertexts, the alphabet used is an unusual one consisting of 29 characters, namely the uppercase letters A-Z and the three punctuations symbols, ',' (comma), '.' (full stop or period) and '-' (dash). The frequency distribution of the individual symbols together with the most common digrams and trigrams are shown in the appendix. Before encryption, all plaintexts characters were changed to upper case and all characters outside this alphabet were deleted.

It is expected that you make use of software tools to help you. All of the tasks are achievable with publicly available tools. It is allowed and encouraged to write your own scripts or programs to simplify some of the repetitive elements but that is not essential. Tools that are recommended are:

- Cryptii: `https://cryptii.com/`
- Dcode: `https://www.dcode.fr/en`

Note that these web tools may try to serve you large amounts of advertising, so you may consider to use them with an ad blocker or in "private" modes.

There are 8 marks available in total for this assignment with part marks as shown. Your answers should be submitted by the due date of **6 February 2023**. Your answers must be delivered through Blackboard in the answer fields provided. Each question also describes what you need to enter. Answers can be added and saved as you go along.

This is an individual assignment. It is acceptable to discuss the general approach with other students but your answers should be your own.

You need to obtain your individual folder of four ciphertext files. The folders, as well as instructions on how to obtain them, are on Blackboard.

# 2 Questions

## Statistical analysis

Using the frequency analysis tool in Dcode or any other tool, find the distribution of individual characters (1-grams), digrams (pairs of letters or 2-grams) and trigrams (3-grams) for each of your ciphertexts. Thus you will end up with four sets of tables, similar to the set in Appendix A.

## 2.1 Caesar analysis (2 marks)

By comparing your tables with the tables in Appendix A, first identify the Caesar cipher. For this you should only need to look at the individual character frequencies, identify the most frequent and check if that gives a plausible shift for all characters. You can use the Cryptii Caesar cipher procedure to decrypt the whole message to check your conclusion.

In the provided answers fields give the number (0, 1, 2 or 3) of your Caesar ciphertext, and the key (shift value) used to encrypt.

## 2.2 Substitution analysis (2 marks)

Use the frequency analysis for characters, digrams and trigrams to identify the simple substitution cipher in your set. Find which ciphertext 3-gram corresponds to the letters THE (upper or lower case). This is not necessarily the most common 3-gram, so check it against the 1-gram and 2-gram distributions as well. There is no need to find a complete key or plaintext.

In the provided answers fields give the number (0, 1, 2 or 3) of your ciphertext, and your assumed encryption of THE.

## 2.3 Vigenère analysis (2 marks)

By comparing your tables with the tables in Appendix A, identify the Vigenère cipher in your set. The index of coincidence tool in Dcode can also be useful. Given that the period is 5, find the plaintext. You should split your ciphertext into five streams and

then apply the same technique that you used to decrypt the Caesar cipher to decrypt each string. You can test your solution using the Vigenère cipher tool in Cryptii.

In the provided answers fields, give the number (0, 1, 2 or 3) of your ciphertext and the key as three letters.

## 2.4 Hill cipher analysis (2 marks)

Find the key for the remaining Hill cipher. This can be challenging. The following process is recommended.

1. Using your digram frequency table, determine candidate digrams (letter pairs) which map to 'th' and to 'he'. These are probably two of the most common of your digrams.

2. Once you have identified likely candidates for two digrams, test them by solving for the key. (This is similar to the known-plaintext attack covered in Lecture 4.) You may need to make a few attempts if the statistics of your individual ciphertexts do not work out nicely. You can test your solution using the Hill Cipher tool in Dcode—note that in Dcode you can only use alphanumeric characters in the alphabet, so if you use this method of testing you will have to change the custom alphabet by substituting, for example, 0 for ',', 1 for '.' and 2 for '-'.

In the provided answers fields give the number (0, 1, 2 or 3) of your Hill ciphertext[1], and the key; write your key as four characters row by row. **If you have not found the key** you can obtain credit for good guesses for the encryption of TH and HE; use the field provided to enter up to three plausible encryptions of TH and HE.

## 2.5 Comment (0 marks)

The final field in the answers allows you to enter any free text if there is anything you want to add. You do not have to write anything.

[1]There is no credit for identifying the Hill cipher since this is already decided once the other ciphertexts are identified.

# A Character, digram and trigram distributions

These tables show the distribution of the single characters, digrams and trigrams taken over the whole source file used to generate the plaintexts.

| Char | % |
|---|---|
| E | 11.93 |
| T | 8.8 |
| A | 7.9 |
| O | 7.52 |
| I | 6.82 |
| H | 6.42 |
| N | 6.23 |
| S | 5.99 |
| R | 5.54 |
| D | 4.36 |
| L | 4.00 |
| U | 2.83 |
| M | 2.64 |
| W | 2.37 |
| Y | 2.24 |
| F | 2.11 |
| C | 2.1 |
| G | 1.89 |
| , | 1.7 |
| . | 1.5 |
| P | 1.47 |
| B | 1.37 |
| V | 0.92 |
| K | 0.75 |
| - | 0.21 |
| X | 0.14 |
| Q | 0.11 |
| J | 0.09 |
| Z | 0.04 |

| Digram | % |
|---|---|
| TH | 2.81 |
| HE | 2.53 |
| IN | 1.64 |
| ER | 1.62 |
| AN | 1.5 |
| HA | 1.34 |
| RE | 1.28 |
| OU | 1.26 |
| AT | 1.08 |
| EN | 1.04 |
| TO | 1.02 |
| IS | 1.02 |
| ES | 1.02 |
| ED | 1.01 |
| ON | 1.00 |
| IT | 1.00 |

| Trigram | % |
|---|---|
| THE | 1.58 |
| YOU | 0.77 |
| AND | 0.74 |
| ING | 0.71 |
| THA | 0.52 |
| HAT | 0.52 |
| ,AN | 0.41 |
| THI | 0.36 |
| HER | 0.35 |
| HIS | 0.32 |
| VER | 0.32 |