

LANGUAGES AND CODING: FREQUENCY ANALYSIS

Please make a selection of texts (newspaper articles, novels, ...) available to the students in each language. If texts are on the internet, they may need them printed out if they are going to do the letter counting by hand. It would be best if students work in groups, splitting the work between them. Results from all groups can then be combined in the spreadsheet (available on the webpage for this day) and bar graphs of letter frequency printed out for the next session.

The issues to discuss can be used to break up the session as you feel appropriate. Discussion could be in the whole group, or in small groups, with views then collected together in a plenary session.

Using as many texts as you have available, find out what the frequency of each letter is in English, French, Spanish and German. Treat letters with accents as if the accent wasn't there. You will need a bar graph of letter frequencies for each language for this afternoon's session.

- If you have access to the internet, use the frequency analyser at http://www.simonsingh.net/The_Black_Chamber/frequencyanalysis.html to take the hard work out of this exercise - and work together on this, splitting up the work between you. **You will need to copy text into Word, and then take out any accents, before putting it into the Frequency Analyser - if you leave accents in, the Analyser will ignore those letters and you will get false results.**
- Alternatively, just count up the number of times each letter appears, splitting the work between you.
- Combine all your results for each language, and put them into the spreadsheet provided to create four bar graphs. Percentages for each letter will appear automatically as you enter the data.
- Print out each graph, so that each group has each one available.
- Use the information to complete the table below.
- You could also make notes on anything that strikes you about the language you are working on that might be useful later when it comes to identifying a coded message in that language.

Issues to discuss

- What do you think about ignoring accents? Is this legitimate, or is a letter with an accent actually a different symbol from the letter without an accent?
- How different are the letter frequencies in the four languages? Are the differences significant or minor? Do you think you'll be able to identify which language a coded message is in from the letter frequencies of from common words or letter combinations?
- Can a different language be thought of as a code, like Julius Caesar's code or Mary Queen of Scots?
- Are people who are good at languages, good at maths, and vice versa? Is this related to the previous question?

	English	French	Spanish	German
Most common letter				
2 nd most common letter				
3 rd most common letter				
4 th most common letter				
5 th most common letter				
Least common letter				
2 nd least common letter				
3 rd least common letter				
4 th least common letter				
5 th least common letter				
One letter words				
Two letter words				
Common three letter words				
Double letters				
Common combinations or words				