**School of Languages, Linguistics and Film Assessed Coursework Coversheet**

For undergraduate (BA) modules coded:

CAT-, COM-, EAL-, FLM-, FRE-, GER-, HSP-, LAN-, LIN-, POR-, RUS-, SML-

**Please read and note the following guidelines:**

1. To assist with anonymous marking, please use your nine-digit student ID number only: do **NOT** use your name anywhere on your coursework.
2. Normally you will be required to submit one electronic copy of coursework via the module’s QMplus area. Most deadlines in this School are set for a Sunday night (23:55). You will be informed by the module organiser of any exceptions to this procedure, either regarding the time or method of submission. It is your responsibility to ensure that you know and meet the submission requirements for each piece of coursework.
3. You must keep a copy of all coursework you have submitted.
4. Extensions to deadlines may ONLY be granted by the Senior Tutor for your department. In order to be granted an extension, you must submit a claim for Extenuating Circumstances BEFORE the coursework deadline. SLLF has an online EC claim form. Details and links to the form can be found on [QMplus School of Languages, Linguistics and Film Landing Page](https://qmplus.qmul.ac.uk/course/view.php?id=4516&section=1).
5. Late submission, without an agreed extension due to extenuating circumstances, will be penalised according to the QMUL regulations relevant to your level of study.
6. Work submitted within 7 DAYS of the deadline will be accepted but subject to a late submission penalty against the marks awarded. The work will be marked normally, and then a late submission penalty of five marks (or 5% of the marks if not marked out of 100) per 24 hour period will then be applied.
7. Work that is more than 7 DAYS late will not be accepted and will not be marked and will receive a mark of ZERO.

**You are reminded that plagiarism, that is copying someone else’s words or ideas without attributing them to that person, is cheating. This is a serious examination offence and at the very least will result in a mark of zero being awarded for this piece of work; it could result in your expulsion from Queen Mary.**

**By handing in this coursework you acknowledge that it represents your own, unaided work and that you have appropriately acknowledged all sources.**

**Please complete the following details:**

Student ID Number:(9-digit number): 190278830

Module CODE and TITLE: LIN6209 CODING FOR LINGUISTS - 2021/22

Title of Coursework: MINI-PROJECT ASSIGNMENT - PART 1

Essay no:

Number of words written:

Module Organiser: Peter McGinty

Seminar Tutor (if applicable):

**Please continue your coursework on the next page**

**Sorry it hasn’t formatted correctly**

**Mini-Project Assignment – Part 1**

My mini-project will be on the frequency analysis of ‘COVID-19’ in online publications from the Daily Mail and the Guardian news outlets.

I have chosen this word as it is very topical in our lives at the moment and there will be plenty of resources for me to explore as it is something that has consumed our society for almost two years. I want to investigate how the frequency of the usage of the word has varied in the past two years, but also on a monthly basis and potentially draw conclusions from this analysis. I will get an even number of articles from both news outlets and will be aiming for at least 20 from each so that I can gain a large data set to examine.

I have chosen the Daily Mail and the Guardian in order to give myself the best attempt of being comprehensive. This is because they are both believed to be on opposite sides of the political spectrum, allowing me to hopefully come close to nullifying the influence that potential political bias could play. The importance of this is paramount as I would like to try to limit contextual factors so that I can take the data for what it is and analysis it accordingly. I cannot simply pick one newspaper as I don’t believe that it will provide me with enough data to draw conclusions from. I will however compare the results from each of these outlets as it would be interesting to see how they differ, but not conclude that this is from potential political bias. I am also able to gain access to these two sources very easily as they are open to view for free to the public online. In order to make them readable for my code, every article that I use for my data, I will put into separate text documents, meaning that I can use ‘open file’ functions.

My application will use two functions to form the base of my investigation. These are the Word\_frequency and Count\_words functions. These of course will be integral to counting how often COVID-19 appears in the articles. I will need to design these so that they can analyse the text documents. I will then need to work out how to create a function that then groups them by month and then year so that I can design different levels to my analysis. I will be running all code that I design in a python IDLE cell so that I make sure that it is able to run and make sure that it carries out its intended function. To check that it works, I will test the code on one article and once I can confirm that it does, I will then use that code on a larger data set.

I will also need to create an average function that allows me to input the data provided by the other functions and sort them by averages over a period of time, such as monthly and also by the News outlet so I can compare findings.

The python cells in Jupyter Notebooks will be useful as they will allow me to create bar charts and graphs displaying my findings. These will be showing the results of the word frequency by month, year and news outlet. This will give my report another medium to express the data that my code has produced – so it can be interpreted in a pictographic form, rather than just numbers and written analysis. After this I can draw conclusions from the data provided.