**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 Student Support Team in SLLF. In order to be granted an extension, you must submit a claim for Extenuating Circumstances BEFORE the coursework deadline. You can claim by using Queen Mary's online form in [MySIS](https://mysis.qmul.ac.uk/urd/sits.urd/run/siw_lgn). Details 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 Queen Mary 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): 19029005

Module CODE and TITLE: LIN6209 – Coding for Linguists

Title of Coursework: Assignment 4 – Part 1

Essay no:

Number of words written: 286

Module Organiser: Peter McGinty

Seminar Tutor (if applicable):

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

**PROJECT PART 1**

**Project title** - ‘Covid Trends’ – Word frequency analysis

**Description on the project** – In this project I will be looking into the word frequency of ‘Covid’ in several different online UK newspaper articles over a period of time (most likely between 2020 – till the present time).

**Outline of how you intend to approach the project:** In this project I intend to source twenty-five articles (five articles for each newspaper company) from a variety of mainstream left- and right-wing UK newspapers, this will include The sun, Daily Mail, The Mirror, Evening standard and the Metro. In order to obtain legitimate text resources and data I will use the data base called Nexis UK which has over 20,000 full text sources. After sourcing my data, I will put each article in txt files so they are ready for the analysis, and I will ensure that they are all within the same folder so that no data be missing from the analysis. From this I will create a code within Juypter notebook and that will count the number of times the word ‘covid’ appears in each article. In order to test that my software works I will write the following: word\_frequency('testMirror1.txt') changing the name to each txt filename and I will also manually count how many times the word appears within the extract to see if it matches with what the code is displaying. Once my data has been collected and I know that the code is running correctly I will then display my result in a bar graph using matplotlib within Jupyter notebook. Lastly, I will write a short analysis underneath on Juypter notebook as to why some articles have a higher word frequency than others.