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.

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): 190242358
Module CODE and TITLE: LIN6209: Coding for Linguists
Title of Coursework: Mini-Project part1
Essay no:
Number of words written: 530
Module Organiser: Peter McGinty
Seminar Tutor (if applicable):

Please continue your coursework on the next page

A project development proposal describing the ideas and objectives you intend to explore and an outline of the software you will build. As well as descriptive text, the document you submit for this assignment can include diagrams, lists, sketches, and so on. It should include the following sections:

1. The project title
2. A description of the project and the question it will explore
3. An outline of how you intend to approach the project:
    i. The data and text resources you will need
    ii. Where you will get free access to sufficient quantities of data?
    iii. The functionality of the application you will create, perhaps with a tentative list of the functions you will build
    iv. How you will you test your software to confirm it works correctly?
    v. The software tools you will use e.g. IDLE & MS Word or Jupyter & MarkDown or something else

250-500 words.

'Brexit' trends

- Frequency analysis of the word 'Brexit' over time in various publications.
    - Can compare three newspapers articles about Brexit from same day across 3 times, (pick significant days of Brexit and two Broadsheets with one Tabloid and get whole newspaper as text file if possible). Label the text files as title of newspaper and year. Then can present results as a comparison.

The frequency analysis of the word 'Brexit' over time and various publications.

The project will focus on three newspapers coverage across three specific dates relating to 'Brexit' and the frequency of the word exploring the question 'Is there a difference in the use of the word 'Brexit' between the vote for Brexit and the leaving of the European Union?'.

The data I need are three articles written by three UK News publications. The three articles that are collected will be published on the three following dates: 23/07/2016, 14/03/2019 and 31/01/2020. This is because on the date 23/07/2016 the referendum was held on and the majority of those who voted chose to leave the European Union; on the 14/03/2019 the UK Government sought permission from the EU to extend Article 50 and agree a later Brexit date (rather than 29/03/2019); and finally, on 31/01/2020 the UK finally left the European Union.

I will get access of the data using the database and research website Nexis. The website Nexis provides news and business information from a range of sources, including UK national and regional newspapers. Due to it being a database it is possible to search groups of sources (for example, UK Broadsheets, Major World Newspapers) or an individual publication. I will filter by

country of publish (UK) and dates (previously stated) and pick the first three results made by The Guardian, The Independent and The Sun. I will separate the articles by saving them as separate text files titled by publication and date. Nexis' coverage dates from the 1980s and is updated daily. I will collect the data by searching the articles online and then creating them as separate text files by both year and publication.

The functionality of the application I plan to create is to be widely used as a tool to help with word frequency comparison. By creating this application, I will have to create functions such as a word count that can recognise word boundaries, including those followed with a space and/or punctuation. I will also have to make sure that I can make the article into a list of words and their occurrence, so I am able to compare the frequency of 'Brexit' across all 9 text files.

To test my function, I will have to execute the code multiple times using Idle and a test text file that I have created myself. In this file the word 'Brexit' will occur a certain number of times to which I will be able to compare the results returned by the code and their actual occurrence with. To do this I will have to count the occurrences of the word myself in the text file beforehand. To overcome any problems with my own counting, I plan to use the list function to return all the words used in the text file with a numerical value. Something which I have previously coded and has passed multiple tests.

For this project I plan to use the software Idle as the main software. This is because I find it easier to read my syntax errors on this software. I plan to use the software Jupyter to create a bar graph to present a comparison of the word frequency across the time and newspapers analysed.