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 - Thursday nights (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. Details 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 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 will be applied. The penalty is five marks (or 5% of the marks if not marked out of 100) per 24 hour period or part thereof.
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): 190861698
Module CODE and TITLE: LIN6209 Coding for Linguistics
Assignment (number or name): Mini Project Part 1
Question (number or title): Project Development Proposal
Number of words written: 557
Module Organiser: Peter McGinty
Seminar Tutor (if applicable): n/a

Please continue your coursework on the next page

Coding Mini Project Part 1:

A project development proposal (500 words) 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:
    1. The data and text resources you will need
    2. Where will you get free access to sufficient quantities of data?
    3. The functionality of the application you will create, perhaps with a tentative list of the functions you will build
    4. How will you test your software to confirm it works correctly?
    5. The software tools you will use e.g. IDLE & MS Word or Jupyter & MarkDown or something else
4. An outline project plan

The project proposal for this assignment will investigate the language used by individuals when reviewing products on Amazon, or more specifically how the language used in Amazon product reviews reflect on the product itself.

Since its founding in 1994, Amazon has become one of the largest (if not the largest) online retailing company in the world. Although it specializes in more than just e-commerce, in this project I will be focusing more on the e-commerce side of the business. Amazon's reputation has been unstable throughout the years, with many claims on its handling of bad products and inhumane working conditions but is still frequently used by many customers around the world. The project that I will be researching will investigate specific products on their website (www.amazon.co.uk) and look at the language used to describe the reviews of the product and whether the reviews are more positive or negative.

To collect the data, I will be using the official Amazon website (www.amazon.co.uk) to search up specific products (clothing, technology, homeware, children's toys etc.). Then when selecting a specific product, I will look at their ratings and reviews and select a range of comments that provide different views on the product to then input into my code to see if the reviews correctly relate to the number of stars that the reviewer has given for the product. When selecting the data that I will need to include a significant number of reviews, ranging from 1-5 stars per product (depending on the choice of product I will check the reviews for).

If the code is successful, this could be used to categorise good and bad reviews for products and may even be used on a broader scope for generalized good and bad comments on businesses and could even be used to hide particularly bad/critical reviews or comments made by the public. To create this code, I will need to categorise terms that are consistently positive and negative and will need an input() function to copy and paste the reviews that I will be investigating into the code. To represent the positive terms, I could need to create a list of positive terms that include 'good' and 'great' as well as phrases such as 'would recommend' and 'I am happy with this product'. The reverse list will be made for negative

terms ‘bad’, ‘unhappy’, ‘not good’ and phrases ‘I would not recommend this product’ and ‘not worth it’.

To test the software, I will have to check without using code whether the reviews will be positive or negative. This could also be done by checking the number of stars given for a review to deem whether it’s more positive or negative. There could be instances where the review is neutral and could be difficult to determine whether the review is more positive or negative overall so a ‘neutral’ or ‘unknown’ result may also have to be included into the code. The software tools that I would use to help create this code will be Jupyter Notebook (to test the code and make sure that it’s functioning correctly) and Microsoft Word (to place all the Amazon reviews into one document that can be easily accessible and so I wouldn’t have to constantly go back onto the Amazon website to look for the product and find the review again if I lose them).