

**COURSEWORK COVER SHEET**

|  |
| --- |
| **Date**: 12 July 2023 |
| **Student number**: 13708652 |
| **Programme of Study**: MSc Social Research |
| **Module**: Masterclass in Social Research |
| **Essay question**: Assignment #1 - Blogpost |
| **Word count (including all footnotes, references and appendices)**: 997 |
| **Disability and dyslexia support**: Do you have an Individual Student Support Agreement with the Birkbeck Disability Office that is relevant to this coursework?  **~~Yes~~** or **No** *(Please delete as appropriate)* |
| **Plagiarism statement**:  Coursework is monitored for plagiarism and if detected may result in disciplinary action. In submitting this coursework, I hereby confirm that I have read Birkbeck's plagiarism guidelines and taken the online tutorial on avoiding plagiarism and on this basis declare that this coursework is free from plagiarism.   * Plagiarism guidelines: <https://www.bbk.ac.uk/student-services/exams/plagiarism-guidelines> * Plagiarism tutorial: <https://www.bbk.ac.uk/student-services/learning-development/preparing-for-study> (scroll down to Citations and Referencing) |

**Text-as-Data**

**Key Highlights**

Words are an important aspect of the social world, with language being the means of communication and as the articulation and codification of important systems. However, the scale of textual data has made its analysis cumbersome, and prior research using texts have primarily been conducted through small-N methods. The contemporary era of information further compounds this issue, with a velocity and volume that makes conventional approaches to studying Big Data near impossible.

Text-as-Data, an increasingly popular approach in computational social science, allows for the analysis of such extensive datasets. It borrows from the literature and methods of computer science and adapts them to social science questions, and as such, the tools it uses are similar to that of Natural Language Processing as well. Analysis is facilitated through languages like Python, which are well-documented and have large, community-driven libraries of models and algorithms that support such work. At the same time, Text-as-Data itself refers to a broad range of such methods, with each one better suited to some aspect of analysis compared to others, and all of them with varying degrees of complexity. On one end of the spectrum, a keywords analysis can be used to identify the occurrence and frequency of certain words in a corpus of data to understand its prevalence. On the other end, unsupervised machine-learning can be used in topic modelling to surface themes and clusters of words based on similarity. These themes may not always be interpretable, but they nonetheless provides a means to quickly sort through a large data set.

**Learnings/Applications in Research**

A large aspect of my learning from this lecture was in understanding the potential feasibility of its use in my dissertation project. Coming from a primarily conventional social research background rooted in qualitative methods, my expectation of text-based analysis was rooted in traditional, small-N methods and a belief that such computational methods from big data were out of reach. I have opted to use keywords analysis in my project to sort through a corpus of Tweets in order to better understand the substance of political conversations on social media, such as the kind of reference links shared during such conversation and the presence of language referring to political activity. Both of these are crucial areas of study in participatory politics and political talk, so Text-as-Data offers an opportunity to study these conversations as they have occurred, furthering the findings derived from survey-driven research.

**More to learn**

I would have liked to learn more about the application of computer science methods to social science activities and the difficulties or blind spots that might occur from the wholesale importation of such methods, and how the measurements of fit and validity used in conventional quantitative methods such as regression are (or are not) applied.

**Discourse Analysis**

**Key Highlights**

Discourse analysis is another approach to studying language. In the Foucauldian tradition, discourse is also about much more than language - it is about social constructions and the way versions of society, events and inner psychological worlds are produced. Crucial to this is a consciousness of power dynamics and the way in which discourse reinforces or reproduces those structures. As such, it is simultaneously anti-realist in denying that there is an external reality awaiting a definitive portrayal, and at the same time constructionist, as it places an emphasis on the versions of reality propounded by members of the social setting being investigated and on the fashioning of that reality through their renditions of it. Meaning isn’t made in isolation, and any analysis must pay attention to the context by which a piece of discourse is made, such as the cultural environment it is situated in and the actors involved and their corresponding roles.

**Learnings/Applications in Research**

If language analysis could be understood as a spectrum of methods running from quantitative to qualitative, then discourse analysis would sit on the qualitative end, opposite from text-as-data methods. By treating discourse as sites in which social worlds are constructed and contested, it challenges meanings and understandings that might otherwise be taken for granted. Emphasis is placed on close reading as opposed to quantitative breadth, with location and the “moment” where the piece of discourse was created regarded as being more important. Discourse analysis was my initial intended approach with my research project, as a substantive study of political conversations on Twitter can contribute to an understanding of the dynamics of sense-making and meaning. I offer an example for consideration:

A screenshot of a video

Description automatically generated

This is a publicly accessible Tweet captured from a keyword search of “20mph”, though I have covered up identifying information. A cursory analysis offers some interesting insights - the Tweet was made by a user that consciously added their title of Doctor in their name; it opens with a question on speed limit policy phrased in a manner that seems to seek relatability; offers personal experience as additional elaboration before concluding with a question of whether they were “n=1”. This was fascinating to me because the combination of their title and the use of science-adjacent language reflected an implicit reproduction of power as it taps on the conventional and social authority granted to “objective science claims” as a means to convince the reader - whether the Tweet itself was or wasn’t objective and scientific is irrelevant, as it is nonetheless informative as to the manner by which persuasion may be attempted in conversations about policy.

This creates an interesting juxtaposition with a Text-as-Data approach, where a sentiment analysis will likely decontextualise the Tweet from the speaker, its intent and the audience involved.

**More to Learn**

Perhaps it is my personal bias towards mixed-methods approaches, but I would have liked to learn more about how qualitative discourse analysis can be applied in conjunction with a quantitative Text-as-Data approach to develop a holistic understanding of political conversations on social media.