Enriching our understanding of historic drought and water scarcity: investigating 200 years of news texts

Helen Baker, Carmen Dayrell and Tony McEnery (Lancaster University, UK)

Drought and the scarcity of water pose substantial challenges to the natural environment within Britain and throughout the world. In order to improve our ability to manage instances of future drought, we would benefit from a greater understanding of the causes and impacts of droughts of the past. Droughts are not necessarily caused by lack of rainfall alone and can be worsened by a number of socio-economic factors such as the rate of water consumption. By increasing our knowledge of attitudes to water usage and scarcity, we can foster more informed public dialogues about issues concerning lack of water.

The UK now has an Environment Agency which monitors instances of drought and acts to reduce their impact but, prior to its establishment, there was no national system which recorded occurrences of drought. Droughts were, however, often mentioned in both national and regional newspapers throughout the nineteenth and twentieth centuries. In this paper, we examine two hundred years of newspaper data, from 1800 to 2014. We aim to quantify and characterise occurrences of drought and water scarcity throughout the United Kingdom and the rest of the world, and explore British public discourses surrounding drought.

One of the initial challenges we faced in gathering the data was identifying an appropriate set of search terms that could be used to find references to droughts. By using query terms which excluded surnames and other irrelevant words, we found that references to drought overwhelmingly related to lack of rain within the nineteenth-century news texts. However, this was not the case when we considered material from the twentieth and twenty-first centuries due to a common tendency to use drought as a metaphor, particularly in journalism relating to sporting achievements (e.g. goal, trophy or scoring droughts) and finance (credit, mortgage droughts). Indeed, in these modern texts we found that the vast majority of references to drought were not presented literally. This meant that we were required to manually clean texts where necessary in order to isolate and exclude metaphorical references to drought.

Another challenge was the fact that newspapers do not discuss drought within the context of Britain only. As we were primarily interested in examining the discourses around drought events in Britain specifically, identifying where the drought happened was a key issue in the analysis. Here, we have utilised an innovative and developing methodological approach which combines corpus linguistics with GIS (Geographic Information Systems) methods, thus adding a useful geographical element to our analyses. In the first instance, this will allow us to visualise collections of data with ease and prove to be of particular benefit when dealing with very large amounts of raw data. By uncovering spatial patterns in types of discourse we are able to 'think geographically' and explore geographically bounded discourses through the examination of the link between textual patterns and geographic references. We have applied a process called concordance geoparsing (Gregory and Hardie 2011, Rupp et al. 2016) to our dataset. This

technique is achieved by means of a number of stages. Firstly, we used corpus linguistics software to extract each occurrence of a search-term and a span of fifty words to the left and right of it. Secondly, these concordances were geo-parsed in order to identify instances of place-names, for which co-ordinates were then found. The results were manually analysed in order to reduce errors. In essence, concordance geo-parsing allows us to find all instances of place-names which co-occur with a search term of our choosing. The co-ordinates we generated were then applied to GIS software for mapping.

Concordance geo-parsing allowed us to divide our data into three broad categories: i) droughts which are reported as occurring in Britain; ii) droughts which are reported as occurring outside of Britain; and iii) references to drought which are mentioned in a more general context, for example in relation to climate change or a breakthrough in research. Each set was analysed by a combination of quantitative and qualitative methods, using standard corpus methods – collocational analysis of selected key words – and closer reading of the concordance lines.

We found that drought and water shortage are mentioned in our dataset in rather vague geographical terms, such as occurring across England or the UK, even though droughts seldom affect an entire country – some affect large parts of a country but others are limited to a relatively small area. Within the nineteenth century, droughts were often mentioned as impeding localised agricultural output. However, writers did associate lack of water with the spread of infectious disease among the population of Britain, particularly in children. In more recent years, drought has taken on a wider frame of reference. Droughts are increasingly mentioned alongside other extreme weather events and in relation to climate change. Writers also show awareness of global issues in referencing drought, particularly those affecting the continent of Africa. Alongside exploring public discourses surrounding drought, we will show how attitudes to water shortage have changed over a period of two hundred years.

References

Gregory, I. N. & Hardie, A. (2011). Visual GISting: bringing together corpus linguistics and Geographical Information Systems. *Literary and Linguistic Computing*, 26(3), 297-314.

Rupp, C. J., Rayson, P., Gregory, I., Hardie, A., Joulain, A. & Hartmann, D. (in press, 2016). Dealing with Heterogeneous Big Data when Geoparsing Historical Corpora. *Paper presented at IEEE Conference on Big Data, Washington, 2014*. Bethesda, MD: IEEE, 80-83.