**Using Geographical Text Analysis to Understand the representation of poverty in UK newspapers**

**Ian Gregory and Laura Paterson**

Analysing large textual corpora in ways that combine an understanding of the broad patterns within the text, while also preserving the detail and nuance is one of the major challenges facing the digital humanities. Geography represents one way of doing this: the challenge is to understand what is being said about particular places and how this varies from place to place. To do this we bring together methods from corpus linguistics, that allow us to summarise and explore large volumes of text, with geographical information science (GISc), that provides the tools to allow us to explore sources geographically. We term his approach Geographical Text Analysis (GTA).

Quantitative geographies of poverty are well known, however the geographies of the perception of poverty are far less well understood. To explore these questions we make use of two large British newspaper corpora from the period 2010-15. The *Guardian* gives us the opinions of left-wing broadsheet newspaper, while the *Daily Mail* provides a contrast from a right-wing tabloid. Together these two newspapers provide us with 380 million words of text from the period. The questions that we explore are: a) what places within the UK do these newspapers associate with poverty; b) what aspects of poverty are they associating with these places; c) how does this vary from place to place, and; d) how the two newspapers differ in their coverage.

To answer these we first use corpus linguistics-based approaches to identify a set of search-terms associated with poverty. Having done this, we then use an approach called concordance geoparsing to identify the place-names that occur near to the search terms. These place-names are allocated to a coordinate. Once the geoparsing has taken place, the overall patterns of poverty in the two newspapers can be mapped. From here, a range of GTA techniques, derived from combining approaches from corpus linguistics and GISc can be used to explore these patterns in more detail and to compare them with each other. We will presents the results of this analysis, while also exploring the wider implications of the advantages and limitations of being able to explore the geographies in large corpora in this way.