N8 CIR Research Case Studies
Susan Fitzmaurice and Seth Mehl

Prof. Susan Fitzmaurice and Dr. Seth Mehl

Susan Fitzmaurice's research centres on the history of the English language, using methodological perspectives provided by historical pragmatics, historical sociolinguistics and computational linguistics. She is particularly interested in the methods and kinds of evidence employed in historical approaches to language study. She focusses on semantic-pragmatic change and the utility of different frameworks for explaining such changes in time and space.

Seth Mehl's research interests lie in corpus semantics, with a focus on methodology relating to various linguistic meanings, from semantics (including semasiological and onomasiological perspectives), to pragmatic and discursive meaning, and grammatical meaning.

Can you give us an overview of the project?

The Linguistic DNA (LDNA) project has been exploring textual meaning by identifying lexical, semantic, and pragmatic patterns in over one billion words, across over 60,000 printed English documents from the 16th and 17th centuries.

The project analyses an extraordinarily large number of lexical co-occurrences – sets of three or more words that tend to occur together in a span of text, such as body, mind, and spirit – in order to model discourse and meaning across time, genres, and authors. LDNA was a collaboration between colleagues at the Universities of Glasgow and Sussex and data specialists in the Humanities Research Institute (HRI).

Did you work with an RSE from the beginning of the project?

Yes, the Digital Humanities Institute Sheffield was a critical partner from the beginning. They contributed to the fundamental design of the research project, a key service they provide to researchers, and they followed through with implementation, all the way to impact.

What was the benefit of working with an RSE? Were there specific tools, software or outcomes you found that an RSE could provide?

In our case, this entailed the development of a sophisticated, bespoke computational linguistic tool, to perform computational tasks that no existing tool could deliver.

Our academic team included linguists, lexicographers, and philologists, who defined the operations we needed to complete – but the research would have been simply impossible without the team of RSEs who could collaborate as partners to translate our needs; help us improve our initial ideas; and make our aims a reality.

What tools and software did you use in the project? Is the software, code and data that you used available for others to reproduce your work?

Our lead RSE, Matthew Groves, with RSE George Ionita, wrote the new bespoke computational linguistic tool using Hadoop, to process big data across multiple virtual machines, and Hive, for warehousing and querying the data outputs, with a web-based interface.

Having worked with an RSE, will it change your approach in the future?

Our work analysing linguistic meaning in large text archives depends on collaboration with RSEs. We rely on their expertise and understanding of humanities research methods, and we value the process of co-production – all of which is necessary to make these projects successful.

N8 CIR
COMPUTATIONALLY INTENSIVE RESEARCH