

Parliamentary Corpora and Research in Political Science and Political History

Luke Blaxill

University of Oxford

<https://www.lukeblaxill.com>

luke.blaxill@hertford.ox.ac.uk

Abstract

This keynote reflects on some of the barriers to digitised parliamentary resources achieving greater impact as research tools in political history and political science. As well as providing a view on researchers' priorities for resource enhancement, I also argue that one of the main challenges for historians and political scientists is simply establishing how to make best use of these datasets through asking new research questions and through understanding and embracing unfamiliar and controversial methods than enable their analysis. I suggest parliamentary resources should be designed and presented to support pioneers trying to publish in often sceptical and traditional fields.

The two decades since the millennium have witnessed a 'data deluge' of digitised sources for research. Scholars working with political texts are amongst the most fortunate beneficiaries, with the digitisation of parliamentary proceedings providing an invaluable resource both for traditional qualitative scholarship and large-scale quantitative text-mining approaches. And yet the impact of the release of digitised datasets in directly inspiring new research in political science and political history has been smaller than might have been hoped for, especially given the publicity generated by generalist exemplar studies (Lansdall-Welfare et al., 2017). Indeed, in the 1970s it was widely believed that computational analysis would come to dominate the humanities and social sciences as the range of resources increased and technology developed (Shorter, 1971). And yet, even in the richly-supported realms of political science and history, remarkably few books and articles have appeared which feature digitised resources such as parliamentary proceedings at their analytical core. This paper looks at some of the reasons for this and suggests some potential solutions.

In political science and political history, researchers tend to be less concerned with language itself, but in language as discourse, and discourse as a means of studying (for example) political change, power, identities, institutions, and cultures. They study parliamentary proceedings with this in mind. This makes the addition of contextual data to parliamentary debates vital to maximising their utility as research tools. Expanding coverage of metadata concerning speakers themselves (e.g. party; seniority; gender); the type of proceedings; who else is in the chamber; speaker interactivity; and other variables, are all extremely welcome. For large scale text mining analyses, classifying topics of debates (which enable large-scale diachronic and international comparisons) has often revolved around the Comparative Agendas Project (Baumgartner et al., 2019) but this has largely focussed on post-1945 data, and many historians and political scientists work on nineteenth cen-

tury proceedings where the Comparative Agendas topic classifications are much less reliable. Other crucial determinants of the 'real meaning' of what is happening in Parliament relate to uncaptured subtleties: speakers often use irony, jokes, vary their tone, make oblique references to current or previous events in the chamber, and respond to unrecorded heckles. All of these escape (or at least partially escape) the textual record. Reconstructing this discursive context helps scholars (and citizens) interpret proceedings more readily.

The challenge of ensuring digitised parliamentary proceedings achieve the maximum research impact in political history and political science runs deeper than resource optimisation. Partly, resource creators and enhancers have been so successful and industrious that the digital provision and enhancement of parliamentary proceedings often runs ahead of the needs of the majority of the history and political science research communities. This means that enhanced digital resources – which allow new research questions to be asked and methods to be used – are published before the community has formulated these new research questions or developed these new methods. The challenge for digitally-inclined political researchers attempting to act as a scholarly vanguard is thus to devise new and interesting research questions that could not have been asked without these datasets and (particularly) to develop analytical methods which will be accepted and impactful in traditional fields such as History, where even rudimentary text mining and linguistic classification are controversial (Guldi and Armitage, 2014; Blaxill, 2020). I will give some thoughts on how parliamentary corpora can be constructed and presented so as to best assist researchers attempting pioneering computer-led analysis in traditional and sceptical fields.

1. References

Baumgartner, F. R., Breunig, C., and Grossman, E. (2019). *Comparative policy agendas: Theory, tools,*

- data*. Oxford University Press.
- Blaxill, L. (2020). *The War of Words: The Language of British Elections, 1880-1914*. Boydell Press.
- Borgman, C. L. (2010). The digital archive: The data deluge arrives in the humanities.
- Guldi, J. and Armitage, D. (2014). *The history manifesto*. Cambridge University Press.
- Lansdall-Welfare, T., Sudhakar, S., Thompson, J., Lewis, J., Team, F. N., and Cristianini, N. (2017). Content analysis of 150 years of british periodicals. *Proceedings of the National Academy of Sciences*, 114(4):E457–E465.
- Partington, A. (2013). Corpus analysis of political language.
- Shorter, E. (1971). *The Historian and the Computer*. Prentice Hall, Englewood Cliffs, N. J.