

Finding Patterns across Multiple Time Series Datasets: Democracy in the Twentieth-century Political Discourses in the United Kingdom, Sweden, and Finland

Risto Turunen (presenter)¹, Hugo Bonin¹, Pasi Ihalainen¹ & Jani Marjanen²

¹University of Jyväskylä, ²University of Helsinki

This paper analyses the contextual variation of nouns and adjectives related to democracy in the United Kingdom, Sweden, and Finland in the twentieth century. We compare parliamentary data (*Hansard*, *Riksdag*, and *Eduskunta*) against press data (UK: *Guardian* and *Times*, Sweden: *Dagens Nyheter* and *Svenska Dagbladet*; Finland: *Helsingin Sanomat* and *Suomen Kuvalehti*). By including both liberal and conservative newspapers as well as parliamentary speeches, our study offers a fresh perspective on the relation between democratic discourses produced by politicians and journalists.

The approach includes visualizing the main similarities and differences in the use of democratic vocabulary between multiple historical time series datasets, as well as applying cross-correlation analysis to automatically find identical patterns between parliament and media or across different nations. The similarity of various word frequency time series charts is evaluated using the Pearson correlation coefficient (PCC), which can vary from -1 to 1. When two time series display simultaneous increases and decreases, the PCC value is nearer to 1 (Derrick & Thomas 2004). The strengths of the PCC are its mathematical simplicity, easy interpretability, and tolerance for noise, while its main limitation is sensitivity to extreme outliers which can be mitigated by using sliding windows to analyze segments of the time series instead of the whole.

Our findings indicate that the cross-correlation is strongest between similar political terms in the same dataset, e.g., the relative frequency of “democracy” and “democratic” over time in a national parliament (in *Hansard* 0.91, *Riksdag* 0.76, and *Eduskunta* 0.65). Another strong set of cross-correlations can be observed when the same political term appears in different datasets from the same country, e.g., the frequency of “democracy” in liberal and conservative press (in the UK 0.87, in Sweden 0.82, and 0.61 in Finland). Transnational correlations of political terms were not as strong as intra-national correlations, but they were clearly evident in the PCC values, e.g., for the frequency of “democracy” they varied from 0.58 to 0.68 between three parliaments under investigation. The shared patterns between parliaments include general increase in the use of “democracy” over time, with notable peaks in the 1930s as a reaction to totalitarianism, around the year 1968 related to the rise of social movements, and in the 1990s, with the expansion of digital communication (Ihalainen et al. 2022). We ensured that our results were not due to intrinsic structural properties of the chosen datasets by calculating the PCC values also for non-political terms, which showed weak or non-existent correlation between political and non-political terms.

Methodologically, our contribution introduces time series methods to the digital humanities, a field which has mostly focused on the manual examination of time series visualizations, with only a few exceptions (Wevers, Gao & Nielbo 2020). From the humanities perspective, we empirically demonstrate the strong linkage between the political discourses in parliament and the press, challenging the notion of parliamentary speech as elite political speech, distinct from a broader society.

References:

- Derrick, T., & Thomas, J. (2004), “Time Series Analysis: The Cross-Correlation Function”. In: N. Stergiou (ed.), *Innovative Analyses of Human Movement*, Human Kinetics Publishers, pp. 189–205.
- Ihalainen, P., Janssen, B., Marjanen, J., & Vaara, V. (2022), “Building and Testing a Comparative Interface on Northwest European Historical Parliamentary Debates: Relative Term Frequency Analysis of British Representative Democracy”. In: *Digital Parliamentary Data in Action*, CEUR Workshop Proceedings, Vol. 3133, pp. 52–68, <http://ceur-ws.org/Vol-3133/paper04.pdf>.
- Wevers, M., Gao, J., & Nielbo, K. (2020), “Tracking the Consumption Junction: Temporal Dependencies between Articles and Advertisements in Dutch Newspapers”, *Digital Humanities Quarterly*, 14(2).