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

Risto Turunen
University of Jyväskylä, Finland
risto.j.turunen@jyu.fi

Pasi Ihalainen
University of Jyväskylä, Finland
risto.j.turunen@jyu.fi

Abstract

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*). While our parliamentary datasets (Ihalainen et al. 2022) encompass several political ideologies simultaneously, the selected newspapers can broadly be categorized into conservative and liberal strands. By including both newspapers with diverse political leanings 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. A value of 1 indicates a perfect positive correlation, where every increase in word frequency in dataset A is matched by a simultaneous increase in dataset B. Conversely, a value of -1 indicates a perfect negative correlation, where every increase in word frequency in dataset A corresponds to a simultaneous decrease in dataset B. The closer the PCC values are to 0, the weaker the relationship between the two variables (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 identifying and addressing outliers before conducting analysis.

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). The

most important finding from a historical viewpoint is the statistically strong cross-correlation between media and parliamentary discourses, with values ranging from 0.55 to 0.76 for the term "democracy". 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 three parliamentary democracies 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 early 1990s with the fall of the Eastern bloc.

Keywords: newspapers, parliamentary speeches, text mining, time series

REFERENCES

Derrick, T., & Thomas, J. (2004). Time series analysis: The cross-correlation function. In N. Stergiou (Ed.), *Innovative Analyses of Human Movement* (pp. 189–205). Human Kinetics Publishers.

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* (pp. 52–68). CEUR Workshop Proceedings, Vol. 3133. 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). http://www.digitalhumanities.org/dhq/vol/14/2/000445/000445.html