

**Cheryl Schonhardt-Bailey****Is there a single 'right' way to study political text?****Article (Accepted version)  
(Unrefereed)****Original citation:**

Schonhardt-Bailey, Cheryl (2012) *Is there a single 'right' way to study political text?* *MethodsNews*, Summer . p. 2.

© 2012 [National Centre for Research Methods](#)

This version available at: <http://eprints.lse.ac.uk/45461/>

Available in LSE Research Online: August 2012

LSE has developed LSE Research Online so that users may access research output of the School. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LSE Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain. You may freely distribute the URL (<http://eprints.lse.ac.uk>) of the LSE Research Online website.

This document is the author's final accepted version of the journal article. There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

# Is there a single ‘right’ way to study political text?

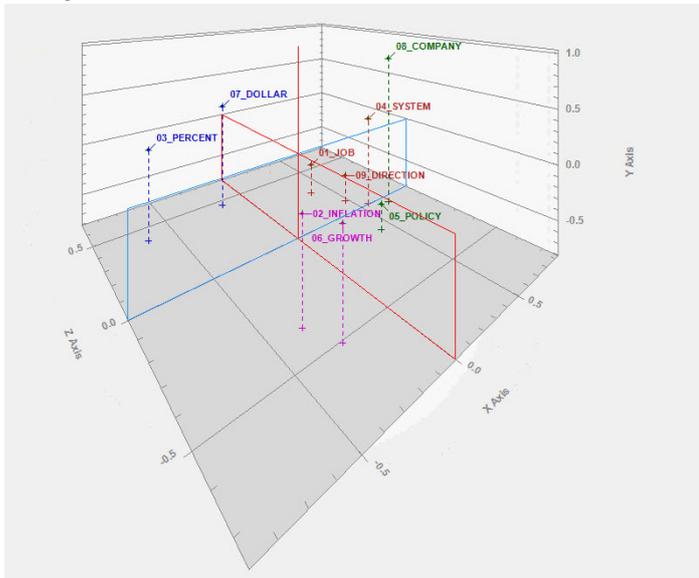
Cheryl Schonhardt-Bailey, London School of Economics and Political Science

**For anyone interested in the empirical analysis of political texts (speeches, committee deliberations, debates, political party manifestos), the simple answer to this question is a resounding “no”. But novices to the growing field of textual analysis might wonder “why not?”**

Why can't political scientists agree on a common toolkit for these texts—something akin to numerical data (regression analysis for interval data; logit or probit models for binary dependent variables; and so on)? There are at least three problems that plague such a toolkit. First, the statistical and theoretical foundations for political textual analysis do not adhere to a single framework, and are thus open to dispute. Second, software packages often fall into one of two categories—proprietary or open-source. The processing methods for the former are invariably opaque while they are usually transparent for the latter.

Social science researchers understandably argue that all algorithms, assumptions and processes of text analysis software should be fully transparent—which implies that they are freely available. There is clearly a tension here between market forces and the development of scientific knowledge (hardly unique to textual analysis), which leads to a third problem: the growing plethora of incompatible textual analysis software which produce fundamentally different types of results.

Using TLab software, a three dimensional correspondence analysis graph of the US House Financial Services Committee, on monetary policy oversight, 1976-2008.



With such obstacles to achieving robust, defensible results from textual analysis, what is the way forward?

One answer to achieving a reasonable threshold of robustness is to ask, do my data look different when I examine them from different perspectives or use different methodological toolkits? If so, one may well have less confidence in the initial approach. If not - if the same fundamental results emerge again and again - the researcher can be fairly certain that she is on solid footing. Looking at data from different perspectives is an increasingly attractive way forward for a number of social scientists<sup>1</sup>. The challenge in this approach can be a rather steep learning curve to acquire the expertise in new software and/or methodologies. Nonetheless, the results can provide one with some degree of internal validation for the research findings.

A second answer might be to seek to validate the findings externally, using entirely different approaches to understanding political texts. For instance, one might employ interviews with the politicians or policymakers who produced the speeches, debates, etc. under investigation, asking them to assess the validity of the findings from the textual analysis. Or one might employ manual coders to read and interpret the raw textual data, thereby providing a check on the initial textual analysis.

Obtaining valid results is, nonetheless, only one of the hurdles confronting the empirical analysis of political texts. More substantively we should ask ourselves, what do we want to know from political texts? If we want to understand the meaning of arguments and debates, we will likely seek to measure how different political actors frame their arguments in order to gain leverage (in issues like abortion or terrorism). Alternatively, we might want to understand the underlying dimensionality of textual data—is it easily captured in a single ideological dimension (usually Left-Right), or are there multiple dimensions required? If the latter, how do we measure and understand political speeches and debates in N dimensions? (The graphs present two examples of committee deliberations in three dimensions<sup>2</sup>) But, the “Holy Grail” of textual analysis is gauging persuasion.

Who, ultimately, is persuaded by the words of political actors, and with what effect? How and why do words matter? Clearly, there is work to be done in this area.

## References

- 1 See, for instance, the papers from the session “Extracting Political Information From Legislative Speeches” of the 5<sup>th</sup> ESRC Research Methods Festival.
- 2 The videos for these graphs may be viewed on the data page of my website <http://personal.lse.ac.uk/schonhar/>

Using Alceste software, a three dimensional correspondence analysis graph of hearings of the US House Financial Services Committee, on monetary policy oversight, 1976-2008.

