Investigating public discourse with Swe-Clarin

**Lars Borin[1], Markus Forsberg[1], Richard Johansson[2],
Tomasz Kosiński[3], Jon Viklund[4]**

[1]Språkbanken, University of Gothenburg,
[2]Computer Science and Engineering, Chalmers University of Technology,
[3]Applied IT, Chalmers University of Technology, [4]Dept. of Literature, Uppsala University
`lars.borin|markus.forsberg@svenska.gu.se,`
`richajo|tomasz.kosinski@chalmers.se, jon.viklund@littvet.uu.se`

Abstract

In the words of Nordmark (2001, 42), "[p]ublic discourse is to be counted among the most remarkable inventions of the early 19th century." The Swe-Clarin center Språkbanken in Gothenburg provides access to almost a billion words of digitized historical Swedish newspapers as well as vast amounts of modern newspapers and digital media, including social media. This creates unprecedented opportunities to investigate public discourse using the sophisticated language tools of the Swe-Clarin research infrastructure. Here we describe two such investigations: (1) the changing attitudes to rhetoric as expressed in public discourse over two centuries; and (2) political issues debated in social media.

1. Changing attitudes to rhetoric

Through its Korp corpus infrastructure, the Swe-Clarin center Språkbanken (the Swedish Language Bank) at the University of Gothenburg provides sophisticated linguistic search tools and online access to over 11 billion words of Swedish texts from various time periods. Notably, there are about one billion words of historical newspapers, about half a billion words of modern newstext, and close to eight billion words of social media (online forums, blogs, etc.). Various visualization options are available, such as trend graphs showing how the use of language expressions changes over time, 'word pictures' showing the typical behavior of words (typical subjects and objects of verbs, typical adjectival modifiers of nouns, etc.), and maps showing geographical locations related to texts or text passages.

Using this infrastructure, we focused on the 19th century newspaper material, and we could see that two conceptual metaphors – *eloquence is fire* and *eloquence is water* – dominated the discourse on rhetoric in this period. This definitely has added to our knowledge of rhetorical history. We also did a preliminary study using 21st century social-media and news text. The two studies together indicate that there have been considerable changes in the attitudes toward eloquence and rhetoric as expressed in Swedish public discourse over the last two centuries.

For further details, see Viklund and Borin (2016).

2. Political discourse in social media

Public discourse has been repeatedly transformed over its long history, its most recent manifestations being the various social media that have emerged only over the last decade or so, complementing or perhaps even supplanting traditional print and broadcast media as the main arena of public discourse and opinion formation. However, there are many questions about public discourse as conducted in social media, questions about the demography and representativity of participation, whether the issues are the same as in traditional media, and whether public opinion formation processes have become fundamentally different as a result.

The work presented here used Swedish tweets collected from Twitter's public streaming API during a narrow time window around two televised Swedish party leader debates in October 2013 and May 2014, before the national elections in September 2014. The aim of the work was to study if issues pertaining to two different political dimensions – *left–right* and *green/alternative/libertarian–traditional/authoritarian/nationalist* (*GAL–TAN*) – were distributed differently in the televised debates and in social media.

The language tools of Språkbanken were used to classify tweets into six different topics reflecting these dimensions, and a spatiotemporal visualization interface was designed allowing researchers to interactively explore the distribution of topics in time and space.

On the one hand, the results of the study indicate that the social-media debate mirrors that in broadcast media, but on the other hand, the interactive visualization shows interesting geographical differences in topic distribution.

For further details, see Borin and Kosiński (2016).

References

Lars Borin and Tomasz Kosiński. 2016. Towards interactive visualization of public discourse in time and space. In *From Digitization to Knowledge 2016*, pages 1–7, Linköping. Linköping Electronic Conference Proceedings.

Dag Nordmark. 2001. Liberalernas segertåg (1830–1858). In Karl-Erik Gustafsson and Per Rydén, editors, *Den svenska pressens historia. II: Åren då allting hände (1830–1897)*, pages 18–125. Ekerlids förlag, Stockholm.

Jon Viklund and Lars Borin. 2016. How can big data help us study rhetorical history? In *Selected Papers from the CLARIN Annual Conference 2015*, pages 79–93, Linköping. Linköping Electronic Conference Proceedings.