Mining Public Discourse for Emerging Dutch Nationalism

Maarten van den Bos, Hermione Giffard

Department of History and Art History, Utrecht University, Netherlands

Asymmetrical Encounters: E-Humanity Approaches to Reference Cultures in Europe, 1815-1992

M.J.A.vandenBos@uu.nl, H.Giffard@uu.nl

ABSTRACT

Historians have often argued that nationalism spread from elite groups to larger populations through public media, yet this has never been empirically proven. In this paper, we propose to study discourse in Dutch public media around events that have been associated with the rise of Dutch nationalism in the current academic discussion. We will use both traditional methods and data mining techniques in order to test the hypotheses that Dutch nationalism grew in strength and changed in focus from the time of the new Dutch constitution in 1848 to the outbreak of the First World War in 1914. This study seeks to develop innovative tools for mining public discourses in large textual repositories.

KEYWORDS: Netherlands, nationalism, text mining, topic modelling, public, discourse

1. Finding Nationalism

The new United Kingdom of the Netherlands was established in 1815. In the 1830s, when the Kingdom fell apart into current-day Belgium, Luxemburg and the Netherlands, its northern region faced the problem of defining a unique national identity. From the adaptation of a liberal constitution in 1848, a new field of 'national politics' became the focus of both public and political debates seeking to define, limit, and challenge the dominant vision of Dutch national identity. (Cf. Sperber, 157, 1984; de Haan, 2003).

Recent Dutch research understands the development of national identity and nation building in the Netherlands primarily through the lens of nationalism studies (de Rooy, 2014). One of the latter's central claims - that public media were crucial in spreading nationalism among the general population – continues to play a fundamental role in academic debate (Anderson, 25, 2006; Zimmer, 2003), although it lacks a foundation in empirical research. We propose to use new digital tools for text mining of big data repositories to study how exactly nationalism was constructed in Dutch public discourse.

By using both traditional methods and data mining tools to study a well-defined corpus, we seek to develop our digital methodology in a way that enables us to scale up our work later to examine even larger and international corpora. Here, we seek to:

1. develop a methodology for mining public discourses digitally;
2. test several hypotheses central to the Dutch historiography about the strength and content of Dutch nationalism over time.

1.1 Historiographical Motivation

In a recent essay, John Breuilly highlighted a 'major deficit' in the field of nationalism studies, both historical and contemporary. These studies, Breuilly states, focus overwhelmingly on either elite discourse or politics. Often, the two concerns are merged into a narrative that stresses the importance of elite discourse in shaping the mind-sets of nationalist politicians. Yet in the evidence that these studies cite – by pointing at high turnouts during nationalist festivals and ceremonies, the popular backing of nationalist politicians, or the high numbers of volunteers for military service in certain stages of a war – one catches glimpses of a larger public audience that not only responded to, but also shaped nationalism.

The problems posed by inferring mass views from those of a limited political elite are particularly urgent within Dutch historiography. We propose to contribute to the historiorahy by studying public media using a bottom-up, data-driven approach. In this way, we hope to explore the historical views of national identity held by different public groups as well as develop a method for new forms of research that will enrich nationalism studies.

1.2 Sources

In the second half of the nineteenth century, especially after the abolishment of the tax on newspapers in 1869, the Dutch media landscape became more and more fragmented. As average incomes went up, the price of papers declined, and literacy rates increased, journals and newspapers became the most important vehicles of new views and opinions

(Van Vree, 2000). Combining this insight with Breuilly's criticism, we have chosen to study public discourse in Dutch public media and will draw our sources from the digitised collections of newspapers and journals of the National Library of the Netherlands. (*Koninklijke Bibliotheek*; kranten.delpher.nl/)

2 Method

To enable a comparison of techniques, our study will begin by selecting from the larger public media corpus several smaller corpora focused around three different, important nationalist events. The corpora will be spread over the time period 1848-1900, allowing us to study the development of Dutch nationalism diachronically. This historical time frame was chosen because public debate about Dutch national identity was most strongly polarized in these years. Each corpus will be analysed using both digital and traditional methods. The intensity and make up of nationalist sentiment at each historical moment will be gauged using a data-driven approach that compares the content and importance of nationalist 'topics' in each case.

2.1 Cases: Events that mobilized and focused Dutch nationalism

We will create a case study corpus around each of several high-profile events in the Netherlands that are strongly associated with an upsurge of Dutch nationalism in the current academic discussion. For each, we will choose a time period around each event that includes about 5000 articles from the digitized collections. Our case studies will include: the 'Aprilbeweging' (1853), the commemoration of the battle for Den Briel at the height of the eighty-years war (1672, commemorated in 1872), and the coronation (1898) of Queen Wilhelmina.

2.2 Computer Analysis

Our method draws on previous studies that have used techniques from computational linguistics to study historical questions (Riddell, 2014; Klein and Eisenstein, 2013). As in these studies, we will use topic modelling to analyse each corpus. We will use the Latent Dirichlet Allocation algorithm to generate groups of words that occur together as well as the distribution of these 'topics' among the articles in each corpus (Blei, Ng, and Jordan, 2003). We will classify each cluster of words so generated by using domain specific knowledge. This work hopes to see whether the known problems inherent in topic modelling, especially in trying to coherently identify and follow a historical subject over time, can be overcome.

By comparing the topics drawn from each corpus, we will try to measure the changing intensity and make-up of Dutch nationalist discourse over time. We expect that economic-nationalist discourse will be strongest around 1850, giving way to political-nationalism by 1900. The percentage of each corpus' words that are represented by nationalist 'topics' will give us a measure of the strength of Dutch nationalism over time.

2.3 Traditional Analysis

The second part of our work will analyse a selection of the corpus (about 100 articles for each event) using traditional historical-interpretative methods. This will enable us to (a) confirm the chosen articles contain nationalist discourse(s) and (b) improve our digital

methods of analysis. In order to integrate comparison into our study, we have chosen a research question that can be answered using either method.

3   Conclusion and Outlook

The current literature on nationalism frequently bases its arguments on vague generalisations and the analysis of elite discourse. We seek to contribute to the literature by leveraging the power of digital text mining to empirically study the emergence of nationalist public discourse in the Netherlands. This project is of methodological interest because we want to develop a method to reliably mine public discourse from public media – both the content of and the degree of public support for, particular debates over time. The aim is to scale up the method described here in order to answer questions about European identity formation through the study of public discourse in larger repositories of historical documents. Unlike this case, which is characterized by a rich historiography from which we drew our hypotheses, we want to be able to confidently identify discourses that are unknown in content. Scaling up the method will also require developing new, advanced methods for corpus selection.

References

Anderson. B., (2006) *Imagined communities. Reflections on the Origins and Spread of Nationalism*. New ed, London.

Blei, D. M., A. Y. Ng, and M. I. Jordan, (2003) 'Latent Dirichlet Allocation.' *Journal of Machine Learning Research 3,* 993–1022.

Breuilly. J., (2012) 'What Does it Mean to Say that Nationalism is 'Popular'?', in *Nationhood from below. Europe in the Long Nineteenth Century*, ed. by Maarten van Ginderachten and Marnix Beyen, pages 23-43, Basingstoke.

Gellner, E., (2008) *Nations and Nationalism*. Cornell University Press.

Haan. I. de, (2003) *Het beginsel van leven en wasdom; De constitutie van de Nederlandse politiek in de negentiende eeuw*. Amsterdam.

Klein, L. and J. Eisenstein, (2013) 'Reading Thomas Jefferson with TopicViz: Towards a Thematic Method for Exploring Large Cultural Archives.' *Scholarly and Research Communication 4.3*. http://www.src-online.ca/index.php/src/article/view/121.

Riddell, A., (2014) 'How to Read 22,198 Journal Articles: Studying the History of German Studies with Topic Models.' In *Distant Readings: Topologies of German Culture in the Long Nineteenth Century*, edited by M. Erlin and L. Tatlock, pages 91-114, Rochester, New York: Camden House.

Rooy, P. de, (2014) *Republiek van Rivaliteiten; Nederland sinds 1813*. 6th ed., Amsterdam.

Sperber. J., (1984) *The European Revolutions, 1848-1851*. Cambridge.

Vree, F. van, (2000) *De politiek van de openbaarheid*. Groningen.

Zimmer, O., (2003) *Nationalism in Europe, 1890-1940*. Basingstoke.