Finding Fascists, Efficiently! Comparing methods for mapping attitudes in Dutch and Belgian historical newspaper corpora (1920-1940)

van den Heede, Pieter

vandenheede@eshcc.eur.nl
Erasmus University, Rotterdam, The Netherlands

van Lange, Milan

m.van.lange@niod.knaw.nl
NIOD Institute for War, Holocaust, and Genocide Studies, Amsterdam, The Netherlands

Futselaar, Ralf

r.futselaar@niod.knaw.nl
Erasmus University, Rotterdam, The Netherlands; NIOD Institute for War, Holocaust, and Genocide Studies, Amsterdam, The Netherlands

There is little positive to be said now about fascist dictators such as Hitler and Mussolini, but they attracted a lot of attention, often positive, in the contemporary European press. In this investigation we use these two men, or rather mentions of their names, as named entities, to evaluate different approaches to investigating attitudes in large-scale digitized historical source collections. Specifically, we will use Belgian and Dutch newspapers to investigate positive attitudes to fascism in the pre-war socialist and Catholic press in these countries.

We aim to apply two methodological approaches to answer the same question on the same dataset. The first of these, word embeddings, have a proven record of effectiveness, and of genuinely revolutionary potential, in historical research (Azarbonyad et al. 2017; Garg et al. 2018; Lange / Futselaar 2019; Wevers / Koolen 2020.) But they are rarely implemented outside of a small community of 'digital humanists'. We believe that the resource-intensive nature of algorithms such as Word2Vec creates a barrier to entry, to academic historians, to students, and especially to our colleagues working in the Global South (Risam / Gil 2022). We will therefore compare the efficacy of word embeddings for answering our historical research question with an alternative, lighter approach based on syntactic dependency parsers.

Our methodological question, then, is whether the alternative method yields comparable historical insights while using less computational power and methodological complexity. We consider this paper an exploration of possible approaches, based on different basic principles of computational methods for text analysis: complex and processing-intensive neural network-based distributional semantics versus a 'minimal computing' approach based on syntax and lexicons (Risam / Gil 2022). To be clear, we fully understand that the output generated by each exercise will be different, but we investigate whether they would nevertheless support a similar interpretation in the context of historical research. Different types of output can play a similar hermeneutic role in historical analysis.

We use these two computer-assisted analyses to investigate positive attitudes towards fascist movements. In both our research countries, the dominant national discourse of pre-war innocence and neutrality has been questioned by historians. They claim that mainstream public opinion and political elites were far more sympathetic to fascist movements than they cared to admit after 1945. Newspaper articles from the interwar years, mentioning fascist leaders, from relatively mainstream newspapers, allow us to test the contention that Dutch and Belgian public and political opinion was less anti-fascist in outlook than was long believed (Vree 1989; Luyten 1996; Lagrou 1999; Boterman 2021). We base our analysis on the socialist newspaper titles *Het Volk/Het Vrije Volk* (The Netherlands) and *De Vooruit* (Belgium) and the Catholic newspapers *De Maasbode* (The Netherlands) and *Het Volk* (Belgium). The newspapers are fully digitized and OCR-ed data are available from the Koninklijke Bibliotheek (The Netherlands) and the Koninklijke Bibliotheek/Bibliothèque Royale (Belgium).

First, we will approach our sources by applying distributional semantics. We use Google's word2vec algorithm to train Word Embedding Models (WEMs) and use their functionality to identify 'nearest neighbors' in semantic space to identify and map attitudes closely related to the named entities of 'Hitler' and 'Mussolini' (Wijffels et al). Second, we will repeat our analysis using a method that relies for its application on less training data and less computer processing power, and is therefore not only simpler, but, as we believe, also more tractable, transparent, and replicable (Risam / Gil 2022). We will use Named Entity Recognition (NER) to identify mentions of both dictators in the digitized historical newspapers, which we know wrote extensively and with strong opinions about both men. Each sentence containing a mention of one of the two is then parsed using a free open-source dependency parser - a technique to examine dependencies between different phrases of a sentence in order to determine its grammatical structure (Benoit / Matsuo n.d.; 'CRAN - Package Udpipe' n.d.). The outputs of the dependency parsers will be used to identify and map contemporary attitudes - and their development over time (Mendelsohn et al. 2020).

Needless to say, there is no external way to validate either approach in the case of diverging outcomes. But in the case of mutually confirmative outcomes, we may well be on the way to establishing a more accessible approach to historical research in large textual datasets. This is especially important because it would help to educate a larger number of historical researchers, widening the critical base for further methodological development.

We try to investigate the needs and necessities of answering a valid and substantive historical research question using digitized historical sources and a computer-assisted approach. We believe that the principles of 'minimal computing' are worth exploration and comparison with established, resource-heavy approaches. In terms of social equity and global access, efficient methods offer a major opportunity for improvement. The ongoing climate emergency also necessitates us to consider the sizeable energy use of heavier computational tools.

Bibliography

Azarbonyad, Hosein / Dehghani, Mostafa / Beelen, Kaspar / Arkut, Alexandra / Marx, Maarten / Kamps, Jaap. (2017): 'Words Are Malleable: Computing Semantic Shifts in Political

and Media Discourse', in: *Proceedings of the 26th ACM International Conference on Information and Knowledge Management.*

Benoit, Kenneth / Matsuo, Akitaka Matsuo.: 'The Spacyr Package for R'. https://cran.r-project.org/web/packages/spacyr/vignettes/using_spacyr.html [07-09-2022].

Boterman, Frits. (2021): *Tussen Utopie en Crisis: Nederland in Het Interbellum 1918-1940* . Amsterdam: Uitgeverij De Arbeiderspers.

Garg, Nikhil / Schiebinger, Londa / Jurafsky, Dan / Zou, James. (2018): 'Word Embeddings Quantify 100 Years of Gender and Ethnic Stereotypes', in: *Proceedings of the National Academy of Sciences* 115, 16. https://doi.org/10.1073/pnas.1720347115.

Lagrou, Pieter. (1999) *The Legacy of Nazi Occupation: Patriotic Memory and National Recovery in Western Europe, 1945–1965* . Cambridge: Cambridge University Press.

Lange, Milan van, / Futselaar, Ralf. (2019): 'Debating Evil: Using Word Embeddings to Analyse Parliamentary Debates on War Criminals in the Netherlands', in: *Contributions to Contemporary History* 59, 1: 140–56.

Luyten, Dirk. (1996): *Ideologisch debat en politieke strijd over het corporatisme tijdens het interbellum in België. Verhandelingen van de KAWLSK van België. Klasse der Letteren.* Brussel: Paleis der Academiën.

Mendelsohn, Julia / Tsvetkov, Yulia / Jurafsky, Dan. (2020): 'A Framework for the Computational Linguistic Analysis of Dehumanization', in: *Frontiers in Artificial Intelligence* 3, 55. https://doi.org/10.3389/frai.2020.00055.

Risam, Roopika / Gil, Alex. (2022): 'Introduction: The Questions of Minimal Computing', in: *Digital Humanities Quarterly* 16, 2. http://www.digitalhumanities.org/dhq/vol/16/2/000646/000646.html.

Vree, Frank van. (1989): *De Nederlandse pers en Duitsland, 1930-1939: Een studie over de vorming van de publieke opinie* . Groningen.

Wevers, Melvin / Koolen, Marijn. (2020): 'Digital Begriffsgeschichte: Tracing Semantic Change Using Word Embeddings', in: *Historical Methods: A Journal of Quantitative and Interdisciplinary History.* 53, 4: 226–43. https://doi.org/10.1080/01615440.2020.1760157.

Wijffels, Jan / BNOSAC / Straka, Milan, Straková, Jana. (2023): 'Udpipe'. R. https://cran.r-project.org/web/packages/udpipe/index.html.

Wijffels, Jan / Fomichev, Max / BNOSAC. (2021): 'Word2vec: Distributed Representations of Words'. R. https://CRAN.R-project.org/package=word2vec.