Neural embeddings for text analysis: a case study in neoliberal discourse

ABSTRACT

This paper examines the notions of neoliberalism and the financialization and marketisation of public life by using computational tools such as sentence embeddings on a novel corpus of neoliberal articles. More specifically, we experimented with distributional semantics along with several NLP techniques and machine learning algorithms in order to extract conceptual dictionaries and "seed" words. Our findings show that sentence embeddings reveal repetitive patterns constructed around the given concepts and highlight the mechanical character of an ideology in its function of providing solutions, policies and constructing stereotypes. This work introduces a novel pipeline for computer-assisted research in discourse analysis and ideology.

Keywords: Discourse Analysis, NLP, Information Extraction, Corpus Linguistics, Ideology, Neoliberalism, Automated Methods, Machine Learning

1. INTRODUCTION

Ideological discourse tends to be stereotypical; it forms constellations of concepts that attract other concepts creating a space that contains its assumptions [1]. Hasan [2] describes, how the influence of late capitalism 'glib speak' on the meanings of concepts such as 'globalization, 'freedom' or 'democracy' is part of a 'semiotic battle' to define reality [2]. As Chaput [3] puts it, this rhetorical circulation of neoliberalism can be found in a variety of phenomena: from the aversion to the welfare state that in neoliberal discourse folds back itself in racist emotional structures [4, 5, 6, 7, 8] to the targeting of any alternative socio-political plan with the aim of accepting capitalism as inevitable and unique.

More specifically, neoliberalism has been described as a new form of capitalism [9], as successive waves of hegemonic regimes [10], restoration of capitalist class power [11], and as a set of individual behaviors and norms [12]. Neoliberalism has been put under the critical lens of gender [13, 14], citizenship [15, 16], biotechnology [17, 18], sexuality [19], work [20], growth [21], environment [22, 23, 24], race [25, 26, 27]. Innset [28] concludes that neoliberalism is not just a set of political propositions but rather a general theory of modernity that now defines most political programs and our very concept of common sense.

In this project, we combine a corpus-based analysis of large data sets with computational methods to identify patterns in the texts around the four nodal concepts of *welfare* and *poverty*. Although each concept represents a multifaceted topic and has been extensively analyzed in the literature, we consider neoliberalism as a useful analytic concept to approach the relationships between the state, the market and society and research the discursive representations of these themes. Our aim is mostly methodological, that is, to examine how Corpus Linguistics and deep learning computational methods on empirical data may be effective in critical approaches. As ideology forms 'through repeated manifestations of a pattern of meaning' [29] computational methods may provide a more objective and quantifiable character to the findings and thus contribute to a better understanding of ideological discursive and stereotypical formations.

2. RELATED WORK

Natural language processing techniques have been widely used in discourse analysis. The quantification of discursive information from large corpora of textual data based on topic models and neural word embeddings have already been widely implemented in social science. In this section we briefly overview neural embeddings on which our pipeline is constructed.

Recent advancements in vector space models have provided analytical tools that help investigate not only the collocational profile of specific words and concepts but also their semantic fields as well as their semantic proximity with other linguistic entities such as other nouns or bigrams. The importance of word vectors for sense disambiguation has been established by several studies [30, 31]. For the purposes of this study, we chose the Word2vec model, a two-layer *neural* network model developed by Mikolov [32] which creates vectors, numerical representations of words by taking into account their context - i.e., a predefined set of words before and after a nodal word. Taking word embeddings one step further, sentence embeddings are often referred to as semantic vector space representations [33]. Embedding the meaning of a sentence into a vector space may result in highly effective applications for natural language tasks, as they enable discourse analysis to exploit a plethora of tools available for computation in vector spaces [34]. In this paper, we experiment with sentence embeddings to detect similarities between sentences in order to establish repetitive patterns and ideas within our corpus.

3. Data and Methodologies

We created a corpus of more than 34,000 articles published in the period from 1978 to 2020. The corpus consists of journalistic articles and blog entries from five well established and widely recognized and neoliberal think-tanks: American Enterprise Institute, Cato Institute, Institute of Economic Affairs and Adam Smith Institute and Cato's project 'Downsizing the Federal Government'.

Table 1: Corpus composition by Institute and number of texts

Think Tank	No of Texts	Unique words
Adam Smith Institute	2820	27103
American Enterprise Institute	24005	144082
CATO Institute	3164	27514
Downsizing the Federal Government	1846	18415
International Economic Affairs	3306	30051

We opt for this specific corpus because think tanks have become critical agents in politics and policy making, although they were formerly supplementing universities and research institutions. Thus, we consider this corpus ideal for the theoretical analysis of ideological and linguistic indicators but also for testing a data-centric approach to the extraction of characteristics to explore the following research question:

- RQ1: Is it possible to detect repetitive argumentative or discursive patterns by measuring the semantic similarity of sentences thus determining stereotypical formations or widespread assumptions within ideologically defined texts? We experiment with sentences vector space models.

3.1 *Word embeddings for semantic fields*

In this study we attempt to highlight the associative and referential meaning of the selected words within the corpus. It should be noted that resulting embeddings are dependent on the data on which they have been trained; therefore, we did not use pretrained models but we custom train the Word2Vec model on our corpus providing a 'window' of 8 words before and after the nodal word and 300 dimensions. Our purpose is to examine the semantic fields of this specific discourse and determine the argumentative environment within which our concepts are discussed. After training our model, we called for a similarity function to examine the vector scores of the selected concepts (Table 2).

Table 2. Concepts vector similarity scores (Word2Vec)

Poverty	**Similarity Score**	***Welfare***	**Similarity Score**
'extreme_poverty'	0.552	*cash_welfare'*	0.505
'absolute_poverty'	0.549	*welfare_program'*	0.494
'child_poverty'	0.542	*welfare_reform'*	0.483
'destitution'	0.501	*tanf'*	0.466
'poverty_rate'	0.500	*welfare_dependency'*	0.465
'relative_poverty'	0.495	*welfare_roll'*	0.446
'material_hardship'	0.483	*benefit'*	0.444
'inequality'	0.480	*social_welfare'*	0.439
'poverty_inequality'	0.478	*entitlement'*	0.431
'poor'	0.476	*welfare_state'*	0.430

As most word embedding algorithms build on the distributional hypothesis [35] where similar contexts imply similar meanings, the similarity between the vector scores of the word *welfare* indicate that *cash_welfare, and welfare_programm* are being negotiated within similar discursive environments. It is also semantically associated with the bigram *welfare_dependecy* and the lemmas *benefit* and *entitlement* (Fig 1). The term poverty bears high similarity scores with the bigrams *extreme_poverty*, *absolute_poverty* and *child_poverty* as well as *inequality*, *reduce_poverty* and *poverty_rate* (Fig. 2).

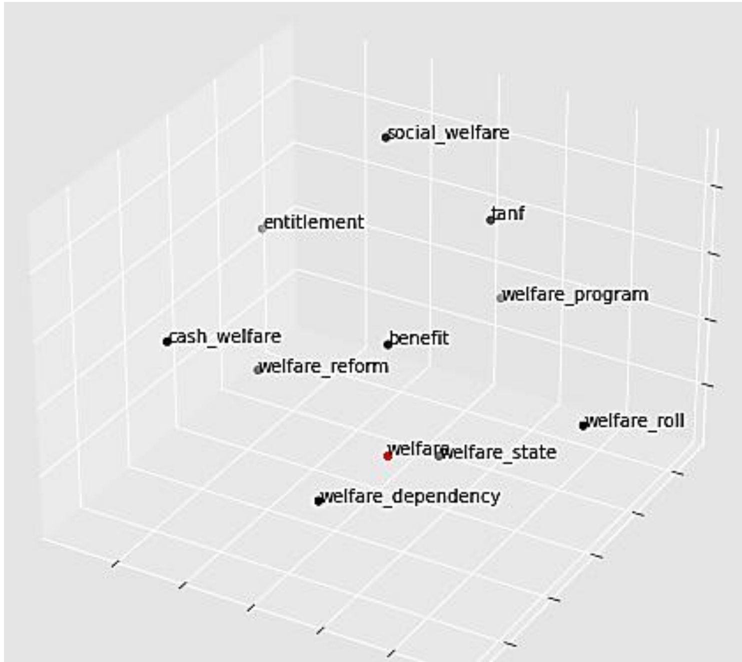

Fig. 1. A three-dimensional representation of the embeddings of *welfare*

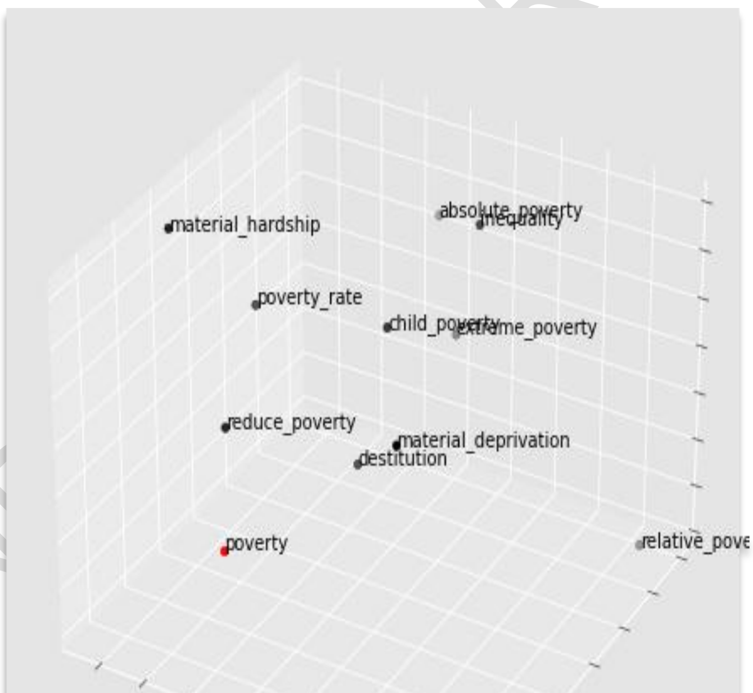

Fig. 2. A three-dimensional representation of the embeddings of *poverty*

3.2 *Representing sentences as numbers*

We expand our analytical query to whole sentences in order to perform a sentence similarity. To this purpose we implement the sentence embedding technique to represent entire sentences and their semantic information as vectors. For this step of the analysis, we extract all the sentences containing the words *welfare* and *poverty* and place them in a data-frame where each line contains an extracted sentence. We preprocess the texts, remove all punctuation and turn the words to lowercase. We then tokenize the sentences, thus creating a list of sentences using Spacy library and its Sentencizer[1] component to allow custom sentence boundary detection. That way we can “feed” the sentences to the embeddings model and we define a function which will return the cosine similarity between the vectors of 2 sentences. We opt for the Universal Sentence Encoder model[2] [36], which takes as input strings and produces as output a fixed dimensional embedding representation of the string. The context-aware word representations are converted to a fixed-length sentence-encoding vector by computing the sum of the representations at each word position. That way the model encodes text into high dimensional vectors that can be used for text classification, semantic similarity, clustering, and other natural language tasks. We use the largest model, which encodes the sentences into vectors of dimension 512 and pass the sentences to the model to generate their embeddings. Finally, we define a *query*, a sentence against which we calculate the similarity, having specified the lowest similarity score at >0.5.

The flexibility of this kind of analysis provides a wide range of experimentation and opportunities to investigate how several concepts are being discussed in a specified field. In the data-frame of the sentences concerning *poverty* we requested the similarities with a phrase taken from our content: *'reduce poverty through aggressive market competition'*. We experimented with a lower threshold of the similarity score to >0.4. Again, the results highlight a significant degree of understanding:

> Sentence = 26387 these are failures of poverty and governmental inadequacy not the market; similarity = 0.44203052
> Sentence = 26802 capitalism and the market proved far better than the state at reducing poverty and raising living standards; similarity = 0.41895163
> Sentence = 29400 the opportunities to fight poverty through the BOP market are endless; similarity = 0.5891824
> Sentence = 41799 we believe that capitalism and free markets lift millions out of poverty; similarity = 0.41657203
> Sentence = 44079 the market economy and the poor but believers in a market economy should not just deliver negative messages about the poverty industry; similarity = 0.46623582
> Sentence = 2741 the answer is free market capitalism; similarity = 0.6215297
> Sentence = 18838 public goods can make markets fail because they tend to be underprovided by the free market; similarity = 0.5479242
> Sentence = 26802 capitalism and the market proved far better than the state at reducing poverty and raising living standards; similarity = 0.5201415
> Sentence = 38182 in Lindsey’s world when the free market doesn’t work government has to step in to give people a helping hand; similarity = 0.51771444
> Sentence = 41373 in the absence of market failures the operation of free markets maximizes social welfare; similarity = 0.5260644
> Sentence = 42640 capitalism and the free market are indispensable; similarity = 0.5307956
> Sentence = 43607 it has become increasingly difficult to make a case for the morality of markets even though free market capitalism has been unequalled in reducing poverty and discrimination and in creating opportunities for social and economic advancement; similarity = 0.5012568

We investigate if the ‘size’ of the welfare state is a repetitive pattern therefore, we insert as a query the phrase “bloated welfare state”, that was already located in our trigrams. Once more we lowered the similarity score to 0.4, as we were not inserting a fully formatted sentence but a phrase. The model generated 316 sentences which successfully detected metaphors as it attributed similarities and transfer of meaning from the word bloated to words like *obese*, *vast*, *costly* and *expanding* as can be seen in the examples below:

[1] https://spacy.io/api/sentencizer
[2] The pre-trained Universal Sentence Encoder is publicly available in Tensorflow-hub (https://www.tensorflow.org/hub)

Sentence = 132 those responses make it sound like people may want a bigger welfare state; similarity = 0.44869676
Sentence = 356 the trump budget is a challenge to congress to start paring back our dangerously bloated welfare state; similarity = 0.48167533
Sentence = 360 social security retirement and medicare should be cut as well but the trump budget provides congress with many good ideas to start paring back the bloated federal welfare state; similarity = 0.41482422
Sentence = 448 the welfare state is so vast and complex that it often works against itself; similarity = 0.44548246
Sentence = 455 alas such contradictions are common in our obese welfare state; similarity = 0.40919337
Sentence = 861 the best we get from government in the welfare state is competent mediocrity; similarity = 0.4273818
Sentence = 1034 instead we provide for their security while they freeride and spend their money on everything else including bloated welfare states; similarity = 0.44298863
Sentence = 1063 republicans point to costly welfare programs such as food stamps while democrats point to the bloated pentagon bureaucracy: similarity = 0.46194437
Sentence = 1142 american taxpayers aren t just asked to support a bloated welfare state at home they re asked to fund free riders in foreign welfare states too; similarity = 0.5074367
Sentence = 1507 geithner' s oped reflects the administration s intransigence in defending the bloated welfare state not any willingness to make serious budget reforms; similarity = 0.4640239
Sentence = 1954 the federal welfare state is expanding rapidly; similarity = 0.5196515

The above examples are only a part of several queries the extracted concepts allowed us to investigate. This technique provides access to locating repeated ideas and *topoi* like the topos of *war* regarding *poverty*, a means to represent *poverty* as an inner enemy of society or the topos of *weighing/burdening down* regrading *government*.

4. CONCLUSIONS AND DISCUSSION

The word embedding model supported the sense disambiguation of the selected words and provided insight into the nature of the discursive environment within which ideas and concepts are being negotiated. We estimate that word embeddings calculated in an ideological discursive space provide access to formulation of the constellations of concepts around nodal points and their latent interchangeability with other terms placed in similar environments. The sentence embeddings, a state-of-the-art method of representing and analyzing text, further highlighted repetitive meanings and whole ideas constructed around specific concepts within a defined ideological space and also highlighted the mechanical character of an ideology in its function of providing solutions, policies and constructing stereotypes. It has been revealed that *markets* are presented as the means to solve poverty and secure even the social state, further establishing the market mechanism into -what may be a reminder of- classic neoliberalism's "spontaneous order" of societal development, as can be also shown by the following examples:

Sentence= 41799 we believe that capitalism and free markets lift millions out of poverty; similarity = 0.561307
Sentence = 42132 in the absence of market failures the operation of free markets maximizes social welfare'; similarity= 0.5260644
Sentence = 43607 it has become increasingly difficult to make a case for the morality of markets even though free market capitalism has been unequalled in reducing poverty and discrimination and in creating opportunities for social and economic advancement

Furthermore, regarding *welfare*, contrary to what might have been expected, our corpus does not promote widespread stereotypes like 'welfare queens', 'scroungers' or any other formations that have been widely implemented by a plethora of media to target specific social groups [37, 38]. Nevertheless, the metaphor of *obesity/bloating* provides the basis for its stereotypical representation: a 'load' of dependent people that must be carried by the energetic, independent members of society.

In general, our analysis has showed that think tanks discourse is far from simply investigative and policy providing as the writers engage strong -occasionally emotional- language which not only contributes to the construction of stereotypical representations it defines the available discursive space [39] for the development of alternative views and for further discursive negotiations over nodal concepts.

This pipeline is providing an unbiased computational approach to a given corpus of textual data for the investigation of ideological nodes and pattern repetitions. As the sentence embeddings showed, it is a promising method for extracting assumptions from a corpus as the similarity of the numerical representations of the sentences do present high degrees of machine learning and 'understanding' repeating ideas and similar discursive formations that can be further analyzed.

REFERENCES

1. Hasan, R. The world in words: Semiotic mediation, tenor and ideology. In J. Webster, editor. Semantic variation: Meaning in society and sociolinguistics. Volume 2 in the collected works of Ruqaiya Hasan, London: Equinox; 2009.
2. Hasan, R. Globalization, literacy and ideology. World Englishes, 2003; 22 [4]:433-448
3. Chaput C. Rhetorical Circulation in Late Capitalism: Neoliberalism and the Overdetermination of Affective Energy. Philosophy & Rhetoric. 2010; 43 [1]: 1-25
4. Tyler I. Revolting subjects: Social abjection and resistance in neoliberal Britain. London: Zed Books; 2015
5. Jensen T. Welfare commonsense, poverty porn and doxosophy'. Sociological Research Online, 2014; 19 [3]: 3.
6. Jensen T., & Tyler I. Benefits broods: The cultural and political crafting of antiwelfare commonsense'. Critical Social Policy. 2015; 35 [4]: 470–491.
7. Hamnett C. Shrinking the welfare state: The structure, geography and impact of British Government benefit cuts. Transactions of the Institute of British Geographers, 2014; 39 [4]: 490–503
8. Jones O. Totally Shameless – How TV Portrays the Working Class. Royal Television Society. 25 November 2013. Accessed 25 February 2020. Available: https://rts.org.uk/article/rts-huw-wheldon-memorial-lecture
9. Brenner N., Peck J. & Theodore N. Variegated Neoliberalization: Geographies, Modalities, Pathways. Global Networks. 2010; 10 [2]:182 – 222.
10. Harvey, D. A Brief History of Neoliberalism. Oxford: Oxford University Press; 2005
11. Duménil G & Lévy D. Capital Resurgent: Roots of the Neoliberal Revolution. Cambridge, MA: Harvard University Press. 2004.
12. Ong A. Neoliberalism as mobile technology. Transactions of the Institute of British Geographers. 2007; 32:3-8.
13. Brown M. Between Neoliberalism and Cultural Conservatism: Spatial Divisions and Multiplications of Hospice Labor in the United States. Gender, Place and Culture: A Journal of Feminist Geography. 2004; 11 [1]: 67–82.
14. MacLeavy J. Reconfiguring work and welfare in the UK's 'New Economy: Regulatory geographies of welfare-to-work at the local level. Gender, Place and Culture, 2011; 18.5: 611–33.
15. Ong A. Neoliberalism as Exception: Mutations in Citizenship and Sovereignty. London: Duke University Press; 2006.
16. Sparke M. A Neoliberal Nexus: Economy, Security and the Biopolitics of Citizenship on the Border. Political Geography. 2006; 25 [2]: 151–80.

17. Birch K. The Neoliberal Underpinnings of the Bioeconomy: The Ideological Discourses and Practices of Economic Competitiveness. Life Sciences, Society and Policy. 2006; 2 [3]: 1.
18. Birch K. Neoliberalising Bioethics: Bias, Enhancement and Economistic Ethics. Life Sciences, Society and Policy. 2008; 4 [2]: 1.
19. Richardson D. Desiring Sameness? The Rise of a Neoliberal Politics of Normalization. Antipode. 2005; 37 [3]: 515–35.
20. Peck J. Political Economies of Scale: Fast Policy, Interscalar Relations, and Neoliberal Workfare. Economic Geography. 2002; 78 [3]: 331–60.
21. Peet R. Geography of Power: Making Global Economic Policy. London: Zed Books; 2007.
22. McCarthy J., Prudham S. Neoliberal Nature and the Nature of Neoliberalism. Geoforum. 2004; 35 [3]: 275–83.
23. Castree N. Neoliberalism and the biophysical environment: a synthesis and evaluation of the research'. Environment and Society: Advances in Research. 2010; 1 [1]: 5-45
24. Klein N. This Changes Everything: Capitalism vs the Climate. New York: Simon & Schuster; 2014
25. Reed A. Marx, Race, and Neoliberalism. New Labor Forum. 2013; 22 [1]: 49–57.
26. Goldberg D.T. The Threat of Race: Reflections on Racial Neoliberalism. Malden: Blackwell; 2009.
27. Haylett C. Illegitimate Subjects? Abject Whites, Neoliberal Modernization, and Middle-Class Multiculturalism. Environment and Planning D: Society and Space. 2001; 19 [3]: 351–70.
28. Innset O. Reinventing Liberalism-The Politics, Philosophy and Economics of Early Neoliberalism [1920-1947]. Springer; 2020
29. Lukin A. Ideology and the text-in-context relation, Functional Linguistics. 2017; 4 [16]. DOI: 10.1186/s40554-017-0050-8
30. Sousa S., Milios E., Berton L. Word sense disambiguation: an evaluation study of semi-supervised approaches with word embeddings. International Joint Conference on Neural Networks. 2020; 1-8, DOI: 10.1109/IJCNN48605.2020.9207225.
31. Iacobacci I., Pilehvar M. T. & Navigl R. Embeddings for Word Sense Disambiguation: An Evaluation Study. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. 2016; 897–907.
32. Mikolov T., Sutskever I., Chen K., Corrado G, & Dean J. Distributed Representations of Words and Phrases and their Compositionality. Advances in Neural Information Processing Systems. 2013; 26: 3111–3119.
33. Iyyer M., Boyd-Graber J & Daume H. Generating sentences from semantic vector space representations. In NIPS Workshop on Learning Semantics. 2014. Accessed 23 April 2021. Available: https://people.cs.umass.edu/~miyyer/pubs/2014_nips_generation.pdf
34. White L., Togneri R., Liu W. & Bennamoun M. How Well Sentence Embeddings Capture Meaning. In Proceedings of the 20th Australasian Document Computing Symposium. 2015; 9: 1–8. DOI: 10.1145/2838931.2838932
35. Harris Z.S. Distributional structure. Word. 1954; 10 [23]:146–162.
36. Cer D., Yang Y., Kong S., Hua N., Limtiaco N., St. John R. et al. Universal Sentence Encoder. ArXiv: 1803.11175; 2018.
37. Cassiman S. Resisting the Neo-liberal Poverty Discourse: On Constructing Deadbeat Dads and Welfare Queens. Sociology Compass. 2008; 2 [5]:1690-1700. DOI: 10.1111/j.1751-9020.2008.00159.x
38. Taylor T. & Bloch K. Welfare Queens and Anchor Babies: A Comparative Study of Stigmatized Mothers in the United States. In M. Vandebeld Giles, editors. Mothering in the Age of Neoliberalism. Demeter Press; 2014
39. Freeden M. Emotions, ideology and politics. Journal of Political Ideologies. 2013; 18 [1]: 1-10