Neural Embeddings for Text Analysis: A Case Study in Neoliberal Discourse

Katerina Mandenaki1*, Catherine Sotirakou1, Constantinos Mourlas1 and Spiros Moschonas1

1Department of Communication and Media Studies, National and Kapodistrian University of Athens, Athens, Greece.

Authors’ contributions

This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.

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 Natural Language Processing (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: Automated methods; corpus linguistics; discourse analysis; ideology; information extraction; machine learning; neoliberalism; natural language processing.

1. INTRODUCTION

Claims about neoliberalism frequently involve assumptions about its discursive life and the societies that form under its influence. Previous work has indicated the hegemonic nature of neoliberal discourse as well as how its assumptions are part of the conditions of the

*Corresponding author: E-mail: mandenaki@yahoo.gr;
formation of particular ‘neoliberalisms’. 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]. Inset [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.

So far, these approaches lack the empirical evidence to map the discursive circulation of neoliberalism. At the same time, the wealth of information resulting from the huge amount of digital texts, combined with the recent affordances of computer science for automatic text analysis, provides new perspectives in discourse analysis, opinion mining, and computational linguistics. Therefore, the utilization of state-of-the-art technologies from the fields of Informatics, Data Science, and Artificial Intelligence (AI) is paving the way to a modern interdisciplinary area that approaches discourse and communication in a novel manner able to understand the use of stereotypes and cognitive models within the context of digital communication.

In this study, we attempt a computational analysis of a large body of extensive textual data in the much-discussed field of neoliberal discourse, and in particular around two nodal concepts of welfare and poverty to explore discursive formations, argumentative patterns and widespread assumptions. 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 to highlight the discursive representations of these themes. We take advantage of a large volume of the data and focus on both the level of the text and the level of the concept in order to define neoliberalism as an ideology. By the same token, our computational approach to neoliberalism brings to light those aspects of its ideology that are mechanical, commonsensical, presupposed. Therefore, our aim is mostly methodological, as we opt to examine how deep learning computational methods on empirical data may be effective in critical approaches. As ideology is formulated ‘through repeated manifestations of patterns 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’.

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>
<thead>
<tr>
<th>Think Tank</th>
<th>No of Texts</th>
<th>Unique words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adam Smith Institute</td>
<td>2820</td>
<td>27103</td>
</tr>
<tr>
<td>American Enterprise Institute</td>
<td>24005</td>
<td>144082</td>
</tr>
<tr>
<td>CATO Institute</td>
<td>3164</td>
<td>27514</td>
</tr>
<tr>
<td>Downsizing the Federal Government</td>
<td>1846</td>
<td>18415</td>
</tr>
<tr>
<td>International Economic Affairs</td>
<td>3306</td>
<td>30051</td>
</tr>
</tbody>
</table>

**Table 1. Corpus composition by institute and number of texts**

<table>
<thead>
<tr>
<th>Poverty</th>
<th>Similarity Score</th>
<th>Welfare</th>
<th>Similarity Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘extreme_poverty’</td>
<td>0.552</td>
<td>cash_welfare’</td>
<td>0.505</td>
</tr>
<tr>
<td>‘absolute_poverty’</td>
<td>0.549</td>
<td>welfare_program’</td>
<td>0.494</td>
</tr>
<tr>
<td>‘child_poverty’</td>
<td>0.542</td>
<td>welfare_reform’</td>
<td>0.483</td>
</tr>
<tr>
<td>‘destitution’</td>
<td>0.501</td>
<td>tan’</td>
<td>0.466</td>
</tr>
<tr>
<td>‘poverty_rate’</td>
<td>0.500</td>
<td>welfare_dependency’</td>
<td>0.465</td>
</tr>
<tr>
<td>‘relative_poverty’</td>
<td>0.495</td>
<td>welfare_roll’</td>
<td>0.446</td>
</tr>
<tr>
<td>‘material_hardship’</td>
<td>0.483</td>
<td>benefit’</td>
<td>0.444</td>
</tr>
<tr>
<td>‘inequality’</td>
<td>0.480</td>
<td>social_welfare’</td>
<td>0.439</td>
</tr>
<tr>
<td>‘poverty_inequality’</td>
<td>0.478</td>
<td>entitlement’</td>
<td>0.431</td>
</tr>
<tr>
<td>‘poor’</td>
<td>0.476</td>
<td>welfare_state’</td>
<td>0.430</td>
</tr>
</tbody>
</table>

**Table 2. Concepts vector similarity scores (Word2Vec)**
As most word embedding algorithms build on the distributional hypothesis [35], where similar contexts imply similar meanings, the proximity of the vector scores between the nodal words and the generated terms supports the sense disambiguation of a word. This function affords the model the capacity to detect ‘transfer’ meanings not only between similar terms (like poverty and material_hardship) but also between literal and metaphorical representations, as we will see in section 3.2. This similarity also implies that in some cases words may be interchangeable. Therefore, the vector similarity 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). Equally, 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 = 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

---

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

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 selected 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. DISCUSSION

The above-mentioned conceptual approach appears to confirm the prominent political analyses of neoliberalism as regards to the selected concepts of poverty and welfare. Regarding poverty our corpus revealed that the general approach of our texts confirms the think tank discursive model which is to offer policies and solutions based on research. What is interesting 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 shown 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 but it defines the available discursive space [39] for the development of alternative views and for further discursive negotiations over nodal concepts.

5. CONCLUSIONS

We created a corpus of blog posts and articles from acknowledged neoliberal think tanks and attempted to approach its assumptions and discursive formations via computational methods to introduce an automated and quantifiable analysis on the empirical data. We opted for a conceptual approach assisted by computational methods inspired by Hasan’s notion of semantic variation. 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 the 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.

This pipeline provides 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 and patterns of meaning from a corpus as the similarity of the numerical representations of the sentences do present high degrees of machine learning and ‘understanding’ discursive formations that can be further analyzed. In the era of big data and digital communication, this methodology can be an asset for any kind of discourse analysis and linguistic research. Future work could further test this methodology in a multilingual context in order to research if the same concepts and meanings are formulated and expressed in different political and geographical frameworks, notably, French and German think tanks. Furthermore, the model could be implemented in different kinds of corpora with the analytical aim to map repetitive patterns and detect specific formulations of meaning.

ACKNOWLEDGEMENTS


COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES


