Using Linguistic Structure to Visualize and Measure Political Concepts

Paul Nulty

Centre for Research in Arts, Social Science and Humanities,
University of Cambridge
pgn26@cam.ac.uk

May 15, 2017

Abstract

Political positions are expressed or implied by the ways in which people use words to construct and relate abstract political concepts and ordinary concepts. This work explores methods for extracting and visualising the lexical environments of abstract political concepts through syntactic parsing of large text corpora. Working in a theoretical framework that treats concepts as cultural entities that can be studied through patterns of lexical behaviour (De Bolla, 2013), I show how natural language processing methods can help researchers to discover and visualise the lexical environments of political concepts. At the level of the sentence, grammatical relation parsing may be used to extract predicates and propositions that compose complex concepts. Beyond the sentence-level, I use a weighted mutual-information measure calculated from long-range term co-occurrences to discover looser conceptual associations that might not occur in a predicating grammatical relation with the central concept. I present examples from historical corpora of the lexical environments in which political concepts operate, and show how these can be compared across time and other variables with interactive tables and network diagrams. Finally, I outline theoretical issues motivating the use linguistic of features for estimating political ideology, and present a preliminary test of this approach.

Quantitative text analysis methods that have been widely applied in political science treat linguistic data as counts of words, phrases, or morphemes within documents. These models ignore the linguistic structure of the text, and fail to capture distinctions of word order, predication, and discourse. Fast and accurate open-source syntactic dependency parsers now make off-the-shelf extraction of linguistic predicates from text relatively accessible. In this paper I explore two ways in which these linguistic features may be useful for studies in political science and the history of ideas:

- Descriptive or qualitative analysis through interactive visualizations

  Keyword search and keyword-in-context views provide a simple digest of the use of particular words in digital text collections. Although not often discussed as a specific method, in practice keyword search and snippet-views of large digital book collections are a widely used in historical and theoretical political research. In the first part of this paper I show how interactive tables and network diagrams may be used to present aggregated counts of linguistic features and word associations, allowing for a descriptive exploration of how concepts are used in the text. This can serve as a level of analysis between a close reading of the whole text collection and a fully automated bag-of-words based classification or topic extraction.

- Linguistic features as parameters for estimation of political ideology

  I motivate and test of the idea of using linguistic features described in the section 3 as parameters in models associating the linguistic features with political dependent variables. Although such models are obviously nothing like the underlying process of understanding the text, if they are a closer approximation to this process than bag-of-words models, then they should be more interpretable and robust.

1 Background: Extracting Conceptual Structure from Text

Advances in computational methods for statistical and linguistic analysis of large digital collections of text allow researchers to investigate latent structure in the shared lexical record at a scale and complexity that was not possible until relatively recently. However, the theoretical promise of distributional or statistical models of meaning has been recognised for some time. In parallel with functional theories of meaning developed in analytic philosophy, linguists began to consider the statistical distribution of morphemes as indications (or even representations) of their meanings. Distributional or statistical semantics as a modern sub-discipline of computational linguistics has roots in early work by the linguist J.R. Firth, who outlined an 'empirical' or 'functional' analysis of meaning:

> [This technique] can be described as a serial contextualization of our facts, context within context, each one being a function, an organ of the bigger context and all contexts finding a place in what may be called the context of culture. It avoids many of the difficulties which arise if meaning is regarded chiefly as a mental relation or historical process. (Firth, 1935)

A pithy restatement of this principle in a later work became a much-cited slogan for the idea: *You shall know a word by the company it keeps*.

Z.S. Harris makes several distinctions between the kinds of contextual patterns that can be used to measure differences in lexical behaviour, including: 'dependence', measured by the tendency for one word to occur close to (within a stateable distance from) another; 'substitutability', a measure of how easily one word may be substituted for another in the same context; and 'selectional preference', lists of words that commonly fill the syntactic argument roles of other words, for example, verbs that often have the same nouns in subject or object position, or nouns that are often modified by the same adjectives.

There is extensive evidence that it is possible to recover complex conceptual structures from large records of linguistic behaviour using computational methods. Studies in cognitive science and neuroscience have used conceptual models derived from statistical corpus analysis to verify the robustness of psychophysical and neuroimaging experiments. In the other

direction, computational linguists have used corpus statistics to replicate conceptual models identified by psychological experiment and neuroimaging. In addition, many natural language processing tasks believed to require conceptual or common-sense knowledge – for example machine translation, analogy solving, question-answering, and natural language inference systems – have been tackled with some success by researchers using distributional semantic models derived from large text corpora.

Irrespective of theoretical motivations, the computational implementations of these methods have much in common. Word, phrase, or document meanings are approximated by deciding on a word-distance window within which to count word co-occurrences or compare the paradigmatic context, and counts of word or context co-occurrence are tabulated into a vector. The vectors can be compared directly to measure word associations, or combined into a matrix to measure the similarity of documents.

Descriptions of distributional semantic methods focus on measuring and evaluating word similarity, but although it is not always explicitly stated, the possibility that these models encode information that could be considered conceptual rather than simply lexical is recognised in the literature. Spatial analogues to meaning such as that of Lund and Burgess are widely cited in cognitive science and validated against human judgements in priming experiments, including in non-linguistic contexts. A key finding is that a single distributional model may be applied to many different tasks that require models of conceptual knowledge (Baroni).

Question answering and information retrieval systems with natural language interfaces exceed human-level performance on many tasks, and neural network language models have been used to label images and perform reasoning over chains of inference. Success on these tasks requires a model of conceptual knowledge. Another strand of research in AI aims to create more explicit knowledge representations for performing inference using typed propositional knowledge in a way that is transparent to the researcher. An example of such a system is ConceptNet, a large cross-linguistic knowledge graph, similar in structure to Wordnet and Framenet, curated from human responses to common-sense or 'practical reasoning' questions.

2 Relation extraction and word association method

The grammatical predicates used in this paper are extracted with a syntactic dependency parser implemented in the SpaCy python package for natural language processing, accessed through the R spacyr package.[1] This parser has been shown to achieve state-of-the-art accuracy on part-of-speech tagging and dependency parsing evaluation datasets (Honnibal, Johnson et al., 2015) and can process thousands of documents per minute on an ordinary system. A dependency parser analyses the grammatical structure of a sentence, establishing relationships between 'head' words and their syntactic modifiers. The table below shows a subset of the linguistic features extracted from a dependency parse of a short sentence.

	token	lemma	pos	dep_rel	targets
32	we	we	NOUN	nsubj	build
33	are	be	VERB	aux	build
34	building	build	VERB	ROOT	build
35	a	a	DET	det	service
36	better	better	ADJ	amod	service
37	Health	health	NOUN	compound	service
38	Service	service	NOUN	dobj	build
39	and	and	CONJ	cc	build
40	providing	provide	VERB	conj	build
41	more	more	ADJ	amod	care
42	care	care	NOUN	dobj	provide
43	for	for	ADP	prep	care
44	those	those	DET	pobj	for
45	in	in	ADP	prep	those
46	need	need	NOUN	pobj	in
47	.	.	PUNCT	punct	build

To explore the characteristic grammatical environments of political terms in whole documents and corpora, we can aggregate counts of these syntactic relations over all of the sentences grouped by a particular variable of interest such as party or annotated ideological position. The next section describes methods for exploring such tables.

In addition to syntactic relations, I make use of word associations derived from co-occurrence of words in adjacent sentences. This method makes a distinction between grammatical and discourse co-occurrence, and does not depend on document divisions alone to count

[1] SpaCy implemtation: https://github.com/explosion/spaCy, spacyr package: https://github.com/kbenoit/spacyr

word co-occurrences in corpora. Considering a document as a list of sentences $s_1, s_2 ... s_N$, one co-occurrence is counted for each word in s_i with each word in $s_i + 1$. That is, for each token in a sentence, we increase by one the co-occurrence count for that word type with the type of every token in the subsequent sentence. We then calculate the overall association between each pair of words that co-occur in this way for the whole corpus, using adjusted Pointwise Mutual Information (PMI). We use the context distribution method of Levy, Goldberg and Dagan (2015) to reduce the impact of very small co-occurrences, with their parameter value of $\alpha = 0.75$.

$$\log \frac{count(cooc(w_1, w_2))}{count(w_1)count(w_2)\alpha}$$

Once this association measure has been calculated for every co-occurring word type in the corpus, it is possible to retrieve a list of the words most associated with a given word of interest. In the next section I describe methods of visualising structure from these lists.

3 Visualizing Structure

The associations detected from a corpus can be viewed as a network, with nodes consisting of words in the vocabulary, and edges representing PMI association scores or counts of syntactic relations exceeding a certain threshold. In this section I outline how interactive network diagrams can be used to explore this data. The figures in this are illustrative and the method should be evaluated using the online interactive prototypes.[2]

Figure 1 shows an image of such a network created from the PMI score associations of the word 'health', extracted from London Times newspaper articles from 1992-1994.

This is an 'neighbourhood' or 'ego' graph of order two, that is, it shows nodes within at most two edges from the focal node — an edge exists between two nodes if their PMI association is a above a given threshold. The graph is drawn using the R *visNetwork* package, using a force-directed algorithm, models the network mechanically as repelling particles connected by springs. The result is that in a graph of suitable density and degree, nodes are spaced apart

[2]Financial Times word association http://52.207.96.220:3838/ft_gui_v2/, London Times word association http://52.207.96.220:3838/apps/times_v1/

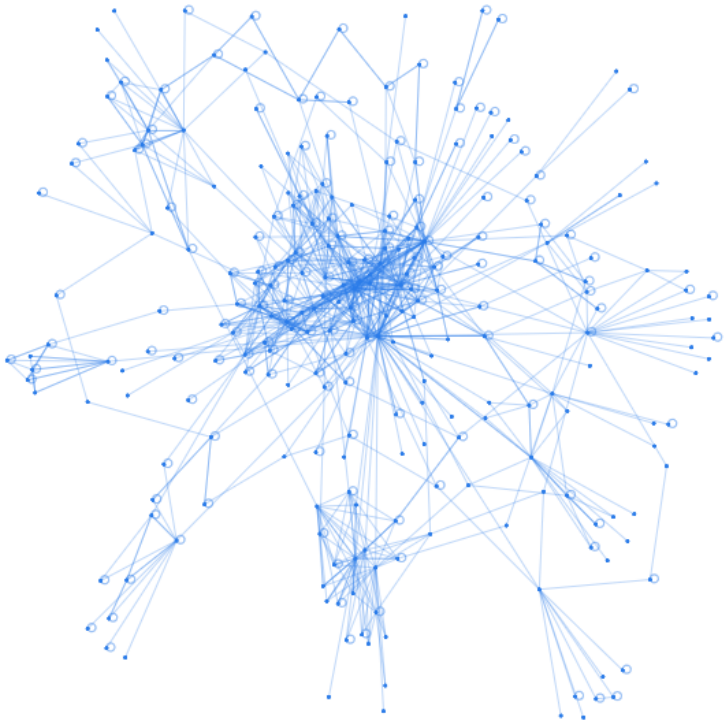

Figure 1: Force-directed network layout of sentence-adjacent word associations

enough to be distinguished, but the edges pull together nodes into clusters that share many relations. In figure 1, the dense cluster at the centre consists of hubs around the terms *patient*, *nhs*, and *infected*, while around the periphery of the graph words related to other topics are connected by bridging terms such as *insurance* and *gymnasium*.

Static images of these networks are of limited use when large enough to show structures larger than a few individual nodes — attempting to label all of the nodes makes them unreadable. If the number of nodes is reduced in order to make the labels legible, then the resulting network is too small to show interesting structure at a large or medium scale. Interpretation or exploration of these semantic networks is therefore best approached through an interactive interface which allows for adjustment in the scale and highlighting of particular neighbourhoods.

Figure 2 is a screenshot of the Shiny interface I have created to explore these graphs.

The visNetwork package implements a drag, pan, and zoom enabled central widget, and this is combined with input boxes for search terms and sliders for setting thresholds or dependent variables. The network in this screenshot is created from PMI associations from Financial Times news articles, with the node *opel/vauxhall* highlighted, which connects the central cluster

around *europe* with a cluster related to car manufacturing.

For demonstration purposes, Shiny apps for the PMI associations created from the FT articles are hosted at here: http://52.207.96.220:3838/ft_gui_v2/

Syntactic relations between words of interest can also be represented in this way, using words as nodes and the type of syntactic relation that holds between them as edges (Figure 3).

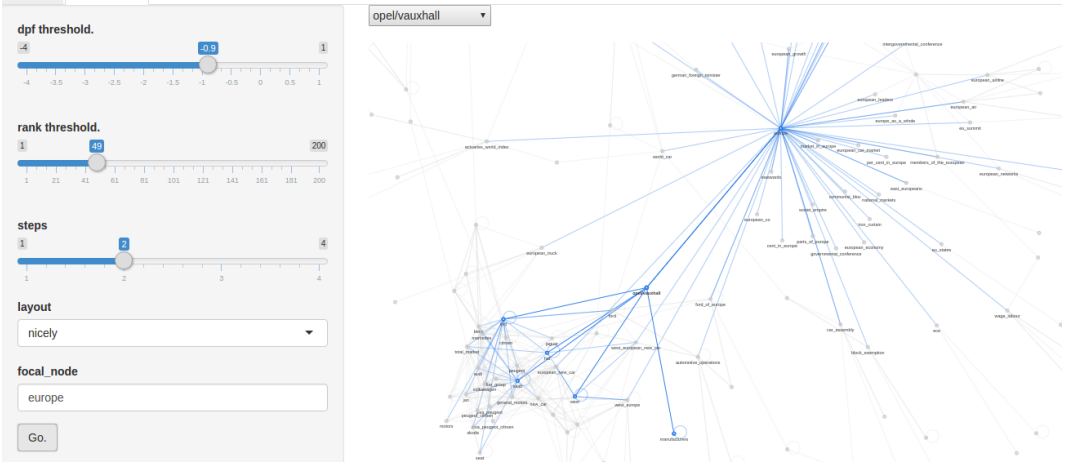

Figure 2: Interface for Shiny application for graph visualisation

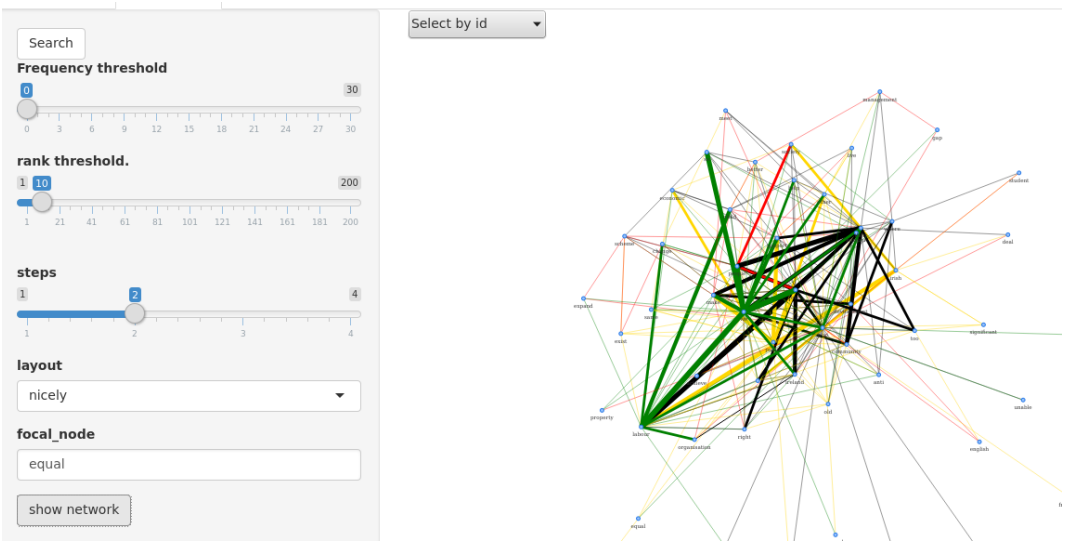

Figure 3: Interface for Shiny application for graph visualisation: the colour of each edge indicates the broad type of syntactic relation that holds between the nodes (verb subject, verb object, modifier, other), and the width of the nodes is determined by the count of that relation in the corpus.

4 Associating linguistic features with political positions

Quantitative social science research usually treats text analysis as a means to an end, where the distribution of words into documents allows for measurement of attention to issues, or estimation of political positions. Document scaling methods solve a practical estimation problems — how similar is each document to the others in the corpus along particular dimensions? (Slapin and Proksch, 2008; Laver, Benoit and Garry, 2003) These estimated positions can then be used in econometric models of the political systems that produced the documents. With some exceptions (Monroe, Colaresi and Quinn, 2008; Sagi, Diermeier and Kaufmann, 2013), the intention is usually *not* to interpret the weights or parameter estimates of the words in the model in order to describe the nature of the relationship between the language used and the resulting political position,:

In this section I motivate and test of the idea of using linguistic features described in the preceding sections as parameters in models associating the linguistic features with political dependent variables. Although such models are obviously nothing like the underlying process of understanding the text, if they are a closer approximation to this process than bag-of-words models, then they should be more interpretable and robust.

4.1 Motivation

The actual means by which a human reader can infer a political position from a linguistic utterance is obviously far too complex to model in an interpretable way, if at all. The huge simplifications assumed in bag-of-words models can produce acceptable results because word counts are often a reasonable indicator of the issue being discussed in a document. The relative difference in issue emphasis reflected by these counts can reflect underlying political positions, even if the model does not capture the actual beliefs or intentions expressed in the text. For example, a unigram bag-of-words model makes no distinction between these sentences:

a) We must reduce the number of nurses in order to increase waiting times.

b) We must increase the number of nurses in order to reduce waiting times.

Nevertheless, the words used in both sentences show that the issue of healthcare is under discussion. This attention to the issue may be what the researcher needs to measure, or it may indicate an ideological position when compared with other texts. In this case, the researcher may choose to use whichever statistical method produces the best empirical classification or regression performance.

The idea that linguistic or semantic information should improve statistical models is implicit in the way that basic linguistic information is commonly used in bag-of-words models in the form of lemmatisation, stopword removal, and phrase detection. The motivation for these steps is that they make models closer approximations of human textual interpretation. Stopwords are removed not because they are uninformative — on the contrary, author identification systems often find them to be the most predictive features — but because we understand that they should have no true direct impact on the dependent variable.

For example, for speeches from a parliament where a large party with an extreme ideological position had a much lower proportion of female representatives than other parties, the pronouns 'she' and 'her' may be highly predictive of left-right position. If the goal is simply to estimate the position of a large number of speeches from this parliament, then the only effect of stopword removal would be to increase the model error, even if cross-validation is used. However, if we wish to generalize the model trained in this parliament by using it to classify in an earlier or later parliament, the gender pronoun estimates may introduce bias. This problem is not confined to bias in prediction of new documents, it is also a source of noise in the in-sample model, insofar as the values of the word parameters vary for reasons that are unconnected with the dependent variable.

There is a practical obstacle to using traditional econometric models for estimating ideology from text: it is possible to fit non-parametric models that assume conditional independence of parameters (such as Naive Bayes and Wordscores) even if there are fewer observations than parameters, but the least-squares method of estimating linear regressions (which don't have this independence assumption) cannot be used when there are more parameters than observations. Again, this problem results from using document divisions as approximations of dependent variables: when an annotator codes a speech for political position, every parameter in the docu-

ment receives the same label regardless of the particular section or sentence it occurs in. Ideally, if researchers are directly interested in the association between language and particular political dimensions, texts should be coded at a more appropriate level (quasi-sentence, or just sentence) (Däubler et al., 2012).

The $p > n$ problem can still occur when using features of the text that separate rather than combine or drop parameters. For example, using syntactic features from the eight-word sentence: "We have fostered a new spirit of enterprise" results in only three parameter occurrences: "We-foster-SUBJ", "new-spirit-MOD" and "spirit-foster-OBJ". However, over the whole document or corpus there may be many more possible types of syntactic relation than observations, in which case a regularized regression model might be used (Zou and Hastie, 2005)

4.2 Experiments

To investigate the association between syntactic relations and political positions I use a corpus created by Benoit et al. (2014) which consists of 18,263 sentences from British Conservative, Labour and Liberal Democrat manifestos for the six general elections held between 1987 and 2010, along with expert coding of each sentence on economic and social dimensions. The expert coders first classified each sentence as referring to economic policy, social policy, or neither. Then, for each sentence relating to economic or social policy, the experts coded the sentence on a five-point ordinal scale from liberal to conservative, with 0 representing 'very liberal' and 5 representing 'very conservative'.

I first focus on the economic coding, retaining only sentences for which four of six experts agree the policy domain is economic. I discard sentences that mention the name of a political party — mentions of party names will presumably be highly associated with political position within this corpus, but this association may not generalize outside of the context of contemporary UK politics. I convert the individual ordinal left-right judgements to a continuous variable by computing the mean of the expert scores for each sentence. This results in a single continuous numerical value for each sentence, on a scale from -2 to 2.

I extract the syntactic dependency type and its target lemma for each token in the sentence, and combine the token, the target, lemma and the dependency type into a single feature.

In an attempt to retain only relations that might explicitly represent beliefs or intentions expressed in the text, only a subset of the possible dependency relations were retained. Pronominal, determining (linking articles with nouns) and rare dependencies types were discarded. To repeat the earlier example, for the sentence ”We have fostered a new spirit of enterprise”, the extracted features are ”We-foster-SUBJ”, ”new-spirit-MOD” and ”spirit-foster-OBJ”.

This process results in 1472 unique feature types and 4133 occurrences in 3801 sentences. I fit a linear regression model with these features as binary parameters and the annotators' mean score for each sentence on the economic left-right scale as the dependent variable. Tables 1 and 2 show a subset of the syntactic features selected by choosing the 200 features with the lowest p-values, and then the 50 of these with the largest estimates in either direction.

	term	estimate	std.error	statistic	p.value
1	asset_rebuild_object	4.34	2.62	1.66	0.10
2	national_on_MOD	3.81	2.61	1.46	0.14
3	more_industry_MOD	3.00	1.60	1.87	0.06
4	economic_mortgage_MOD	2.69	1.63	1.65	0.10
5	economic_make_MOD	2.62	1.10	2.38	0.02
6	economy_low_REL	2.54	1.11	2.28	0.02
7	basic_pension_MOD	2.40	1.33	1.81	0.07
8	new_help_MOD	2.40	1.31	1.84	0.07
9	at_extra_REL	2.23	1.24	1.80	0.07
10	low_any_MOD	2.18	1.34	1.64	0.10
11	better_we_REL	2.13	1.41	1.51	0.13
12	rule_fiscal_object	2.08	1.42	1.46	0.14
13	new_cent_MOD	2.03	1.03	1.98	0.05
14	people_who_object	2.03	1.20	1.69	0.09
15	share_hold_object	2.00	1.18	1.69	0.09
16	same_we_MOD	1.99	1.22	1.64	0.10
17	top_50_MOD	1.98	1.25	1.58	0.11
18	our_how_REL	1.97	1.18	1.67	0.09
19	we_how_SUBJ	1.91	0.94	2.03	0.04
20	local_power_MOD	1.82	1.17	1.56	0.12
21	worth_1_MOD	1.77	1.13	1.56	0.12
22	better_we_MOD	1.69	0.90	1.87	0.06
23	our_economy_REL	1.68	1.04	1.61	0.11
24	benefit_new_object	1.65	0.91	1.80	0.07
25	rate_tax_object	1.62	0.99	1.64	0.10
26	credit_new_object	1.55	0.80	1.93	0.05
27	high_service_MOD	1.53	0.83	1.85	0.06
28	business_we_SUBJ	1.53	0.92	1.65	0.10
29	be_way_REL	1.48	0.84	1.76	0.08
30	we_british_SUBJ	1.46	0.78	1.88	0.06
31	opportunity_have_object	1.42	0.96	1.48	0.14
32	new_building_MOD	1.40	0.71	1.97	0.05
33	we_look_SUBJ	1.39	0.86	1.62	0.11
34	increase_big_object	1.36	0.78	1.75	0.08
35	stake_direct_object	1.35	0.75	1.79	0.07
36	scheme_introduce_object	1.33	0.88	1.52	0.13
37	tax_reduce_object	1.32	0.91	1.44	0.15
38	national_prosperity_MOD	1.24	0.81	1.54	0.12
39	stable_low_MOD	1.24	0.87	1.43	0.15
40	unnecessary_reduce_MOD	1.23	0.71	1.72	0.08
41	inflation_low_object	1.22	0.85	1.44	0.15
42	school_extra_SUBJ	1.22	0.71	1.72	0.09
43	we_create_SUBJ	1.22	0.86	1.42	0.16
44	sustainable_we_MOD	1.21	0.80	1.52	0.13
45	we_easy_SUBJ	1.21	0.72	1.69	0.09
46	personal_we_MOD	1.21	0.65	1.88	0.06
47	band_10p_object	1.21	0.72	1.68	0.09
48	social_not_MOD	1.21	0.75	1.60	0.11
49	private_competition_MOD	1.17	0.81	1.44	0.15
50	private_contribute_MOD	1.16	0.65	1.77	0.08

Table 1: Syntactic relations associated with economic right-wing scores.

	term	estimate	std.error	statistic	p.value
1	free_16_MOD	-4.47	3.08	-1.45	0.15
2	spend_national_object	-4.38	2.87	-1.53	0.13
3	we_rebuild_SUBJ	-3.63	1.62	-2.24	0.03
4	we_get_SUBJ	-2.90	1.74	-1.67	0.10
5	at_this_REL	-2.62	1.77	-1.48	0.14
6	major_programme_MOD	-2.42	1.19	-2.04	0.04
7	great_choice_MOD	-2.40	1.42	-1.69	0.09
8	we_begin_SUBJ	-2.37	0.84	-2.83	0.00
9	enterprise_help_object	-2.25	1.12	-2.01	0.04
10	care_personal_object	-2.25	1.31	-1.72	0.08
11	public_our_MOD	-2.24	1.20	-1.87	0.06
12	first_house_MOD	-2.24	1.31	-1.72	0.09
13	extra_job_MOD	-2.24	1.37	-1.63	0.10
14	more_be_REL	-2.23	0.80	-2.78	0.01
15	rate_per_SUBJ	-2.09	0.87	-2.40	0.02
16	industry_privatise_object	-2.00	1.14	-1.76	0.08
17	low_review_MOD	-2.00	1.31	-1.53	0.13
18	major_europe_MOD	-1.95	1.23	-1.58	0.11
19	public_get_MOD	-1.95	0.78	-2.51	0.01
20	who_choose_SUBJ	-1.92	0.92	-2.08	0.04
21	low_sound_MOD	-1.90	1.08	-1.76	0.08
22	we_responsibility_SUBJ	-1.87	0.87	-2.14	0.03
23	other_pay_MOD	-1.81	0.92	-1.96	0.05
24	operative_co_REL	-1.79	1.02	-1.76	0.08
25	dramatic_increase_MOD	-1.79	0.78	-2.28	0.02
26	we_replace_SUBJ	-1.78	0.85	-2.08	0.04
27	we_want_SUBJ	-1.77	0.82	-2.17	0.03
28	people_pay_SUBJ	-1.74	1.10	-1.59	0.11
29	high_tax_MOD	-1.72	0.83	-2.08	0.04
30	service_we_object	-1.69	0.79	-2.14	0.03
31	we_capital_SUBJ	-1.68	0.84	-2.01	0.04
32	direct_we_MOD	-1.64	0.93	-1.77	0.08
33	we_at_SUBJ	-1.59	0.73	-2.18	0.03
34	we_all_SUBJ	-1.58	0.78	-2.03	0.04
35	we_financial_SUBJ	-1.55	0.85	-1.83	0.07
36	annuity_you_object	-1.53	0.65	-2.35	0.02
37	more_flexible_REL	-1.53	0.65	-2.35	0.02
38	people_work_object	-1.51	0.92	-1.63	0.10
39	recognised_qualification_MOD	-1.51	0.84	-1.80	0.07
40	fiscal_we_MOD	-1.48	0.91	-1.63	0.10
41	growth_economic_object	-1.47	0.92	-1.60	0.11
42	we_burden_SUBJ	-1.44	0.55	-2.62	0.01
43	we_between_SUBJ	-1.43	0.65	-2.19	0.03
44	deal_poor_object	-1.43	0.65	-2.19	0.03
45	market_open_object	-1.43	0.96	-1.49	0.14
46	rate_top_object	-1.42	0.72	-1.98	0.05
47	extra_at_MOD	-1.42	0.95	-1.50	0.13
48	do_our_REL	-1.42	0.88	-1.61	0.11
49	half_be_SUBJ	-1.42	0.88	-1.60	0.11
50	red_have_MOD	-1.41	0.86	-1.64	0.10

Table 2: Syntactic relations associated with economic left-wing scores.

5 Discussion and future work

While many of the features most associated with right-wing scores do capture well-known economic conservative tropes — *fiscal rule, have opportunity, reduce tax, low inflation*, others simply reflect the issues being discussed rather than a clear right-wing agenda or *tax rate, local power, extra school*. The same is broadly true of Table 2: the syntactic features reflect the issues under discussion in economic discourse in much the same way as ordinary word features would.

There are two possible avenues for developing this work, depending on whether it is possible to discover better interpretable features by changing the model specification.

An alternative model of the association between text features and political position is to treat each possible class of syntactic relation as a categorical variable, and the lemmas linked by the relation as possible values of the variable. That is, the model would have in total four independent variables: SUBJ, OBJ, MOD, and REL, and for the sentence "We raise taxes", the values would be SUBJ:"We-raise" and OBJ:"raise-taxes". It might also be useful to explore features that encode the full subject-relation-object triple.

Another remaining question is whether the linguistic features do in fact provide more robust generalization when tested out-of-sample. Cross-validation experiments with elastic net regression failed to achieve greater accuracy than bag-of-words models. To fully test this more annotated text is required, as the data is too sparse when segmented into heterogeneous segments.

References

Benoit, Kenneth, Drew Conway, Benjamin E Lauderdale, Michael Laver and Slava Mikhaylov. 2014. "CROWD-SOURCED TEXT ANALYSIS: REPRODUCIBLE AND AGILE PRODUCTION OF POLITICAL DATA.".

Däubler, Thomas, Kenneth Benoit, Slava Mikhaylov and Michael Laver. 2012. "Natural sentences as valid units for coded political texts." *British Journal of Political Science* 42(04):937–951.

De Bolla, Peter. 2013. *The architecture of concepts: The historical formation of human rights*. Fordham Press.

Firth, John Rupert. 1935. "THE TECHNIQUE OF SEMANTICS." *Transactions of the philological society* 34(1):36–73.

Honnibal, Matthew, Mark Johnson et al. 2015. An Improved Non-monotonic Transition System for Dependency Parsing. In *EMNLP*. pp. 1373–1378.

Laver, Michael, Kenneth Benoit and John Garry. 2003. "Extracting policy positions from political texts using words as data." *American Political Science Review* 97(02):311–331.

Levy, Omer, Yoav Goldberg and Ido Dagan. 2015. "Improving distributional similarity with lessons learned from word embeddings." *Transactions of the Association for Computational Linguistics* 3:211–225.

Monroe, Burt L, Michael P Colaresi and Kevin M Quinn. 2008. "Fightin'words: Lexical feature selection and evaluation for identifying the content of political conflict." *Political Analysis* 16(4):372–403.

Sagi, Eyal, Daniel Diermeier and Stefan Kaufmann. 2013. "Identifying Issue Frames in Text." *PLoS one* 8(7):e69185.

Slapin, Jonathan B and Sven-Oliver Proksch. 2008. "A scaling model for estimating time-series party positions from texts." *American Journal of Political Science* 52(3):705–722.

Zou, Hui and Trevor Hastie. 2005. "Regularization and variable selection via the elastic net." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67(2):301–320.