

Finding positions in parliamentary text

Graeme Hirst
University of Toronto

Supported by the Natural Sciences and
Engineering Research Council of Canada



NSERC
CRSNG

Semantic search

- Search Web or document collection by meaning, position, argument, ideology.
 - Automatically answer complex questions.
 - Automatically summarize and synthesize information from multiple documents.

Not just search but research

Examples

- Find evidence that ...
 - ... Norway is capable of developing WMD.
 - ... society is too tolerant of drunk drivers.
 - ... the Prime Minister is doing a great job.

A text might answer a question without any intent by its author that it do so.

Examples

- Find editorials in German newspapers that support the Antwerp debt-reduction plan.
- How do supporters and opponents of the *Cabbage Abatement Act* justify their positions?
- Please summarize the arguments for and against the proposed Eglinton tramway.

Examples

- Find arguments for and against Obamacare, whose frame is ...
 - ... economic benefits / disadvantages;
 - ... what the role of government should be;
 - ... whether health care is a human right.
- What frames are commonly used in arguments for and against Obamacare?

Components of this idea 1

- Language analysis:
 - Semantic interpretation at the sentence level.
 - Find *discourse relations* between sentences.
 - Determine structure of argumentation.

Components of this idea 2

- Task-oriented:
 - Sentiment analysis – positive or negative view?
 - Recognition of *position* or *stance*.
 - Recognition of *framing* and *ideology*.

The role of vocabulary

- In a debate on some topic, where in the language is ideology apparent? Expect ...
 - ... vocabulary relates to only topic of debate;
 - ... ideology is apparent only at sentence-level and text-meaning level.
- In fact, different ideological frameworks lead to different vocabulary for same topic.

Past research 1

- Thomas *et al* 2006 on U.S. Congress:
 - Does speaker support or oppose legislation?
 - Automatic classification based on words used and on (dis-)agreement with others.
 - Accuracy = 70% (baseline = 58%).
- Greene 2007:
 - Add syntactic relations, get 74%.

Past research 2

- Diermeier *et al* 2007:
 - Automatically classify U.S. senators as liberal or conservative by vocabulary only.
 - Accuracy on “extreme” senators = 94%.*
Accuracy on “moderate” senators = 52%.
 - Found some easy shibboleths:
gay → liberal, *homosexual* → conservative.

*Artificially high due to overlap of training and test data.

Past research 3

- *Yu et al 2008*:
 - Automatically classify U.S. Congress members as Democrat or Republican by vocabulary only.
 - Accuracy on House of Reps = 80%.
Accuracy on Senate = 86%.

Past research 4

- DEFT (Défi Fouille de Textes) 2009:
 - Automatic classification of MEPs' speeches by party (for five largest parties).
 - Mediocre results.

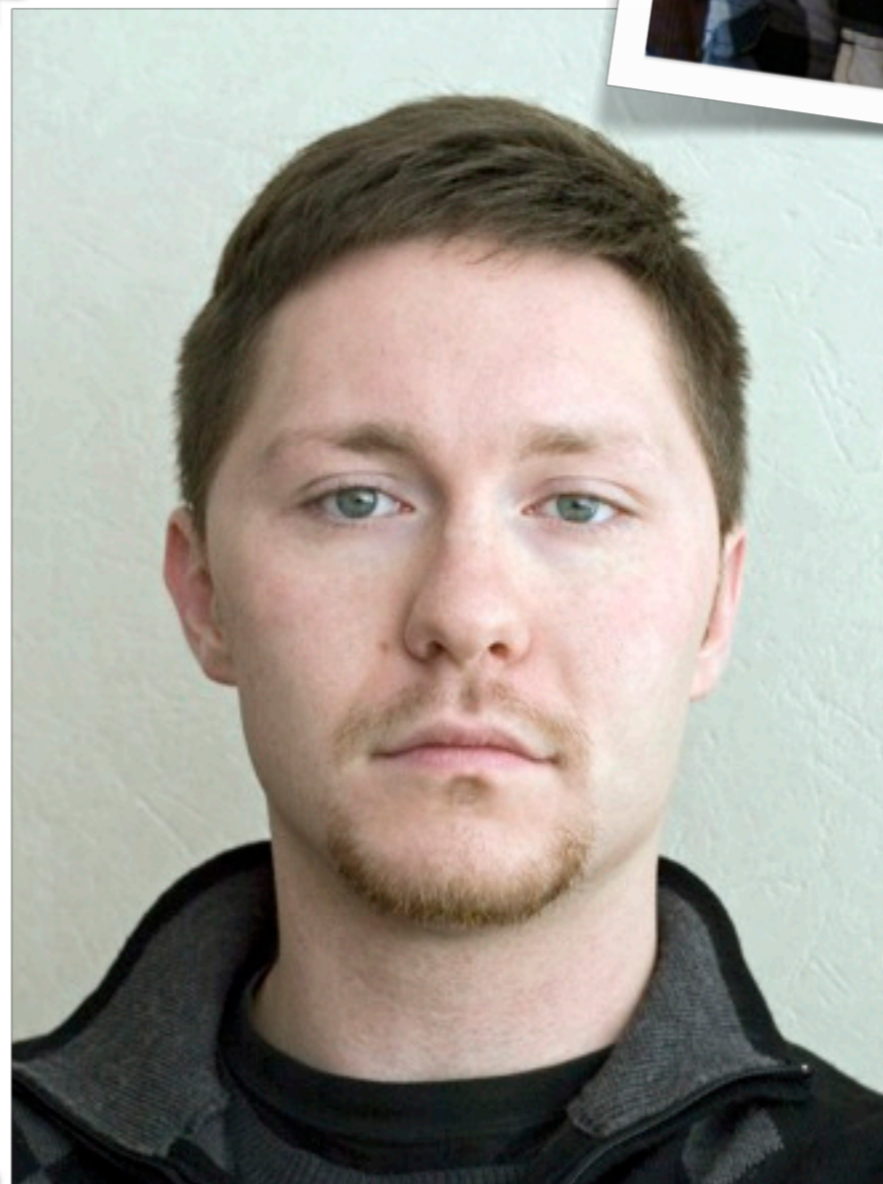
Automatic Classification of Political Speech by Party Membership

Party status as a confound
for lexical methods



Magali Boizot-Roche

Yaroslav Riabinin



Jory Graham

Research question

- Can we identify the liberal / conservative ideology of Canadian MPs by their choice of words?
 - ▶ Party membership as proxy for ideology
- Results: Yes, but ...

Background details

- Parliament includes Prime Minister and Cabinet
- Strong party discipline; no coalitions
- Debates (GOV) and Oral Question Period (OQP)
- Language = {English, French}
 - ▶ All text is translated into other language

Framework

- Data = *Hansard* = Canadian Parliamentary proceedings (House of Commons only)
- 36th Parliament: Liberal government, two conservative opposition parties (ignore other parties for now)
- For each member, all utterances → bag of words

Framework

- Classification method = Support-vector machine, 5-fold cross-validation
- Features = words, *tf-idf* weighting
- High-frequency words retained or discarded

Results

Accuracy of liberal / conservative classification (%)
(frequent words retained)

	OQP +GOV	OQP	GOV
English	83.8	96.9	83.3
French	83.2	89.5	86.0

Majority baseline = 65.5%

But discriminating vocabulary is **not** ideological ...

	Liberal	Conservative
1	hon	prime
2	member	why
3	we	liberal
4	opposite	solicitor
5	quebec	farmers
6	housing	finance
7	bloc	he
8	reform	liberals
9	québécois	hrdc
10	women	banks

... and is barely ideological even when frequent features are removed

Liberal

- *congratulate, excellent, progress, established, inform, improve, assist, developing, promote*
- *housing, violence, humanitarian, youth, society, technology*

Conservative

- *justify, resign, failed, admit, refusing, mismanage*
- *taxpayer, dollar, millions, paying, premiums*

We have not classified by party ideology
— liberal or conservative

We have classified by party status
— government or opposition
— defender or attacker

Testing this hypothesis (1)

- “Ideological” classifier should be robust
 - ▶ when party status is different in test data from training data ...
 - ▶ ... or is varied in training and test data
- 39th Parliament
 - ▶ Party status swapped from 36th:
Conservative government, liberal opposition

Results

Train on 36th, test on 39th

Accuracy of liberal / conservative classification (%)
(frequent words retained)

	OQP +GOV	OQP	GOV
English	44.9	43.3	44.6
French	45.7	46.1	47.0

Majority baseline = 55.8%

Results

Train on 39th, test on 36th

Accuracy of liberal / conservative classification (%)
(frequent words retained)

	OQP +GOV	OQP	GOV
English	36.8	34.5	36.2
French	35.2	51.1	33.5

Majority baseline = 65.5%

Results

Train and test on members of both

Accuracy of liberal / conservative classification (%)
(frequent words retained)

	OQP +GOV	OQP	GOV
English	62.0	66.9	66.1
French	63.0	63.0	63.0

Majority baseline = 64.0%

Testing this hypothesis (2)

- Discriminating vocabulary of “ideological” classifier should not change when party status does

Discriminating vocabulary changes sides

36th Parliament

39th Parliament

	Liberal government	Conservative opposition	Liberal opposition	Conservative government
1	hon	prime	conservatives	bloc
2	member	why	prime	liberals
3	we	liberal	conservative	senate
4	opposite	solicitor	immigration	violent
5	quebec	farmers	mulroney	we
6	housing	finance	kyoto	québécois
7	bloc	he	admit	greenhouse
8	reform	liberals	minority	ndp
9	québécois	hrdc	promise	corruption
10	women	banks	her	member

Testing this hypothesis (3)

- “Ideological” classifier should degrade / fail
 - ▶ if ideologies are muddied in data
- 36th Parliament
 - ▶ Liberal government *versus* conservative opposition parties and left-wing parties

Class is consistent in status
but inconsistent in ideology

Results

Accuracy of liberal / heterogeneous-others classification (%)
(frequent words retained)

	OQP	GOV
English	95.6	82.6

Majority baseline = 51.5%

Emotion words predict party status

- Happy liberals, dour conservatives?
- *Result:* Positive and negative emotion words discriminate parties in OQP with 73–81% accuracy, in GOV with 55–80% accuracy
- *But:* Positive emotion words characterize governing party, negative characterizes opposition

Discussion (1)

- Language of attack and defence dominates and confounds ideology in Canadian Parliament.
- Confound could occur in any attempted ideological classification of speech

Discussion (2)

- Better results than prior U.S. research
 - ▶ But for the wrong reason!
- Reflection of Canada / U.S. differences?
 - ▶ Congress more substantive, separate from Executive
 - ▶ Weaker party discipline in U.S.

Testing this in Europe

- European Parliament has ...
 - ▶ No government and opposition *per se*
 - ▶ Many parties with wide range of ideologies
 - ▶ Primarily ideological debate (we surmise)



Colin Morris

Framework

- **Data = Proceedings of European Parliament**
(Thanks to Maarten Marx)
 - ▶ For each member, all utterances → set of bags of words (we experimented with bag size)
- **Language = English**
- **Parties = {left, right} or [left .. right]**

Framework

- Classification method = Support-vector machine, 5-fold cross-validation
- Features = words, *tf-idf* weighting



European United Left – Nordic Green Left

Socialist, eco-socialist, communist
Speakers: 104

Progressive Alliance of Socialists and Democrats

Social democrats, centre-left
Speakers: 446

The Greens – European Free Alliance

Green, regionalist
Speakers: 114

Alliance of Liberals and Democrats

Centrist, liberal
Speakers: 195

European People’s Party

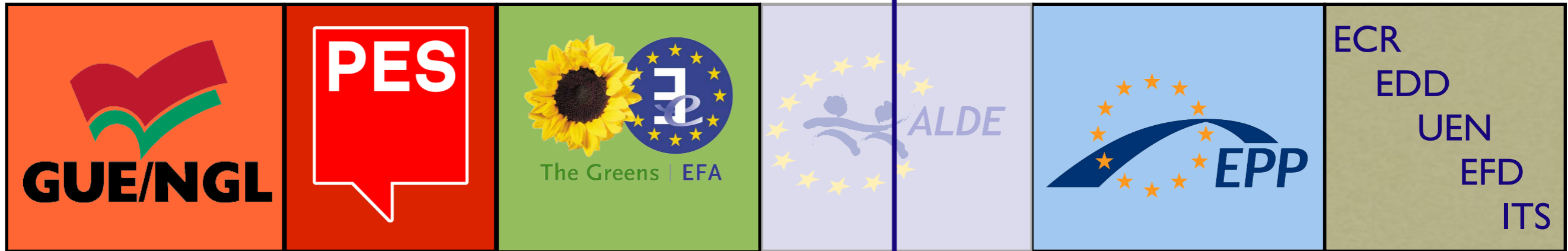
Christian democrats, centre-right
Speakers: 571

Small right-wing groups

Conservative, eurosceptic, nationalist, far-right
Speakers: 231

Left
←

Right
→



European United Left – Nordic Green Left
Socialist, eco-socialist, communist
Speakers: 104

Progressive Alliance of Socialists and Democrats
Social democrats, centre-left
Speakers: 446

The Greens – European Free Alliance
Green, regionalist
Speakers: 114

Alliance of Liberals and Democrats
Centrist, liberal
Speakers: 195

European People's Party
Christian democrats, centre-right
Speakers: 571

Small right-wing groups
Conservative, eurosceptic, nationalist, far-right
Speakers: 231

Results

Accuracy of left / right classification
(large bags of words)

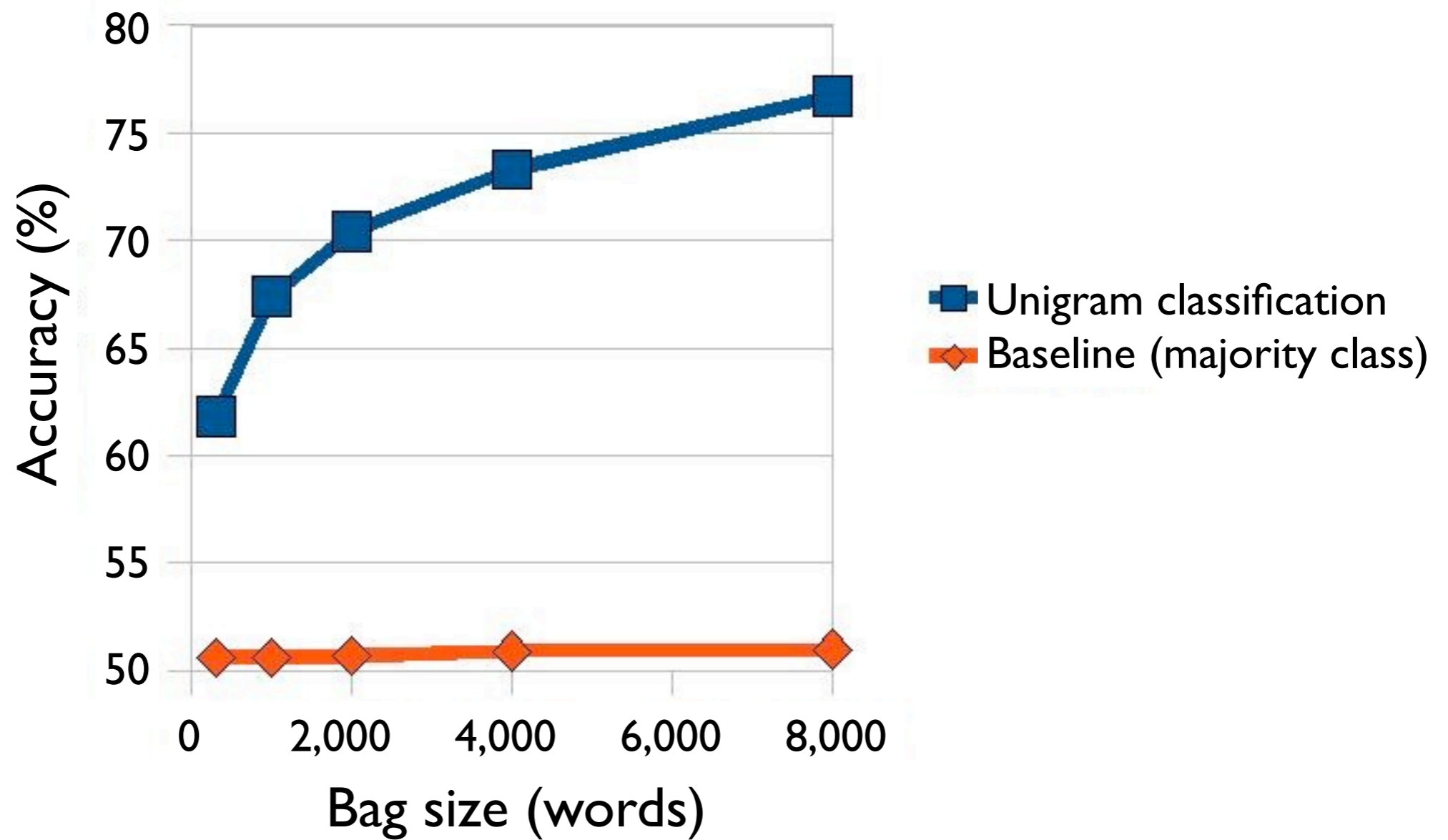
78.5%

Majority baseline = 50.5%

Discriminating vocabulary is more ideological than in Canadian data ...

	Left	Right
1	socialist(s)	subsidiarity
2	unions	christian
3	pse	competitiveness
4	employees	strasbourg
5	greens	healthy
6	scotland	prosperity
7	gender	democrats
8	equality	competitive
9	supports	communist
10	myself	truth

Left / right classification accuracy by size of bags of words





European United Left – Nordic Green Left
Socialist, eco-socialist, communist



Progressive Alliance of Socialists and Democrats
Social democrats, centre-left



The Greens – European Free Alliance
Green, regionalist



Alliance of Liberals and Democrats
Centrist, liberal



European People's Party
Christian democrats, centre-right



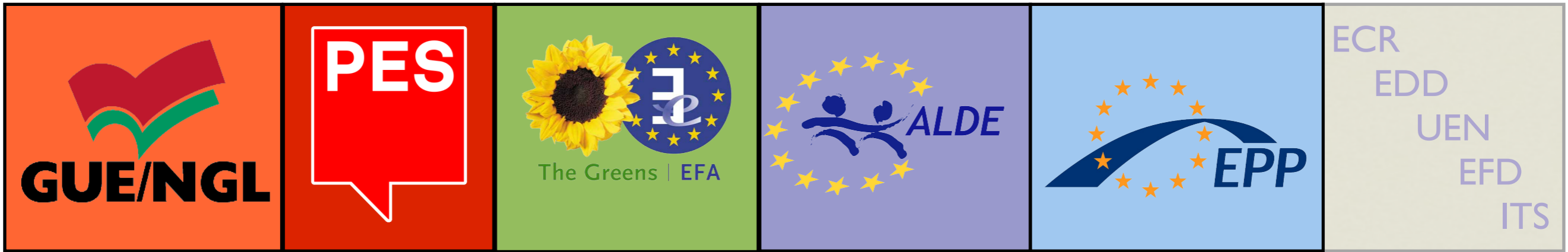
Small right-wing groups
Conservative, eurosceptic, nationalist, far-right

Results

Accuracy of 5-way classification by party
(large bags of words)

61.8%

Most-frequent baseline = 38.5%



- | | | | | | |
|---|--|--|---|---|--|
| <ul style="list-style-type: none"> 1. confederal 2. nordic 3. neoliberal 4. profits 5. militarisation 7. privatisation 10. ngl 11. deregulation 12. multinationals | <ul style="list-style-type: none"> 1. social wages 2. social employment 3. post-capitalist 4. epp wage 5. military inequality 6. gender refugees 12. 54 researchers 15. 59 poverty 21. berlusconi | <ul style="list-style-type: none"> 1. greens childcare 2. 25 decongratulation 3. scotland congratulation 4. basque minorities 5. ale 44. repressio 6. vert 67. lobbyists 8. nuclear unionists 11. stateless 19. organic | <ul style="list-style-type: none"> 1. liberal anti-racism 2. epp radioactivity 3. liberals 4. democratic 5. alde. chemical 10. accountability 11. needless 13. shameful | <ul style="list-style-type: none"> 1. christian arrears 2. subsidiarity 3. conservatives 4. democrats 5. communism 6. democrats 12. competitiveness 20. sports 26. competitive | <ul style="list-style-type: none"> 37. taxpayers 44. proves 45. wrong 47. hezbollah 58. farmers 66. faith 69. euthanasia 73. taxes |
|---|--|--|---|---|--|

Confusion matrix for 5-party classification

(in 1350 large bags of words, approx equal group representation)

Classified as ...

		Classified as ...					Total
		NGL	PSE	Greens	ALDE	PPE	
True affiliation	NGL	204	17	36	9	10	276
	PSE	16	136	20	34	71	277
	Greens	20	25	153	30	16	244
	ALDE	3	39	14	170	50	276
	PPE	3	65	9	41	159	277
	Total	246	282	232	284	306	1350
Accuracy (%)	73.9	49.0	62.7	61.5	57.4	61.8	

Confusion matrix for 5-party classification

(in 1350 large bags of words, approx equal group representation)

Classified as ...

		Classified as ...					Total
		NGL	PSE	Greens	ALDE	PPE	
True affiliation	NGL	204	17	36	9	10	276
	PSE	16	136	20	34	71	277
	Greens	20	25	153	30	16	244
	ALDE	3	39	14	170	50	276
	PPE	3	65	9	41	159	277
	Total	246	282	232	284	306	1350
Accuracy (%)		73.9	49.0	62.7	61.5	57.4	61.8

Confusion matrix for 5-party classification

(in 1350 large bags of words, approx equal group representation)

Classified as ...

		Classified as ...					Total
		NGL	PSE	Greens	ALDE	PPE	
True affiliation	NGL	204	17	36	9	10	276
	PSE	16	136	20	34	71	277
	Greens	20	25	153	30	16	244
	ALDE	3	39	14	170	50	276
	PPE	3	65	9	41	159	277
	Total	246	282	232	284	306	1350
Accuracy (%)		73.9	49.0	62.7	61.5	57.4	61.8

Discussion (1)

- Far better results than DEFT 2009
 - ▶ Bag-of-word size matters
- 5-way party classification results almost as good as 2-way left / right results
 - ▶ Left / right ideology not a good model for Europe
 - ▶ Separate dimension of euroscepticism; Green sentiment somewhat orthogonal too

Hix, Simon; Noury, Abdul G; and Roland, Gérard. *Democratic Politics in the European Parliament*. Cambridge University Press, 2007.

Discussion (2)

- Parties tend to talk about themselves and their special issues
 - ▶ NGL: *nordic, wages, unemployment, profits*
PSE: *socialists, pse, wholehearted, congratulations*
Greens: *greens, organic, nuclear, toxic, ecological*
ALDE: *liberal, accountability, shameful, shame, trials*
PPE: *christian, conservatives, moral, faith, conscience*
- Hint of attack in ALDE (censorious words) and defence in PSE (words of felicitation)

Discussion (3)

- Subjectively, PSE and PPE have “less coherent” vocabularies
 - ▶ Perhaps the cause of their greater confounding

What next?

- All the above used only bags of words.
 - Effects found even though topics are the same.
- No consideration of word order or context.
- Can we do better with smarter methods?
 - Word pairs
 - Relationships between words
 - Use of semantic markup, political mashups

Ideological framing

- Want to automatically find cues to *ideological framing* of political discourse.
 - Quantifiable semantic characteristics thereof.
- Look at relationships between sentences.
 - Find structure of arguments and of discourse.
 - Find *enthymemes* and implicit assumptions.

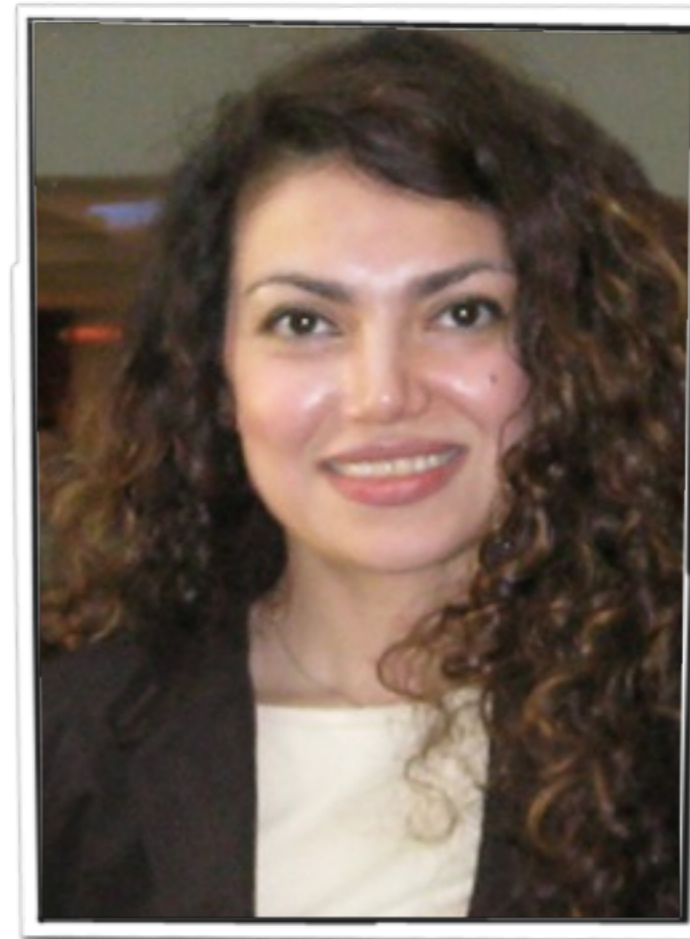
Digging into

Linked Parliamentary Data

- Canadian Parliament, from 1867 (English and French) with metadata.
- U.K. Parliament, from 1803, XML format.
- Dutch Parliament, from 1814, XML format, rich annotations.



CHRISTOPHER COCHRANE
Political Science



NONA NADERI
Computer Science

Conclusion

- Automatic analysis of political and opinionated text.
 - Positions, arguments, ideologies.
- Use in study of ideology and opinion.
- Use in searches; for automatic question-answering, summaries and syntheses.

The End