Measuring Agenda Change in Political Discourse

Azadeh Nematzadeh Norbert Chan

Abstract

Applications of automated textual analysis to the research on issue framing often assume the stability of issue frames. However, both theories and empirical studies suggest the instability of issue frames, where political actors shift from one frame to another in response to temporal events such as interest group advocacy, electoral activities, and institutional reforms. As a result, the classification of political texts is prone to the problem of concept drift where the underlying properties of the target concept change over time. Using machine learning, we modify the existing approach to text classification to accommodate change in issue ownership and issue frame.

Applying machine learning for concept drift analysis, we develop a computational approach to detect and track temporal changes of (1) issue ownership and (2) issue frames. In the former application, we identify the critical words that come to characterize the policy focus of each political group at different time steps. In the latter application, we focus on the rate and pattern of frame turnover in the parliamentary discourse on controversial policy issues. We compiled and applied a diachronic corpus from the published chamber remarks of the National Assembly of France in the period 2003 to 2010 and (to be specified), respectively. Our results indicate that term preferences change temporally and the asymmetry is potentially meaningful in terms of policy deliberations. They also support our future efforts to account for concept drift in analyzing long-running political deliberations.

1 Introduction

Broadly defined, concept drift refers to the temporal change of the properties of a given concept or attribute. Concept drift represents a major source of challenge in the field of textual analysis. For example, while spam consistently refers to illegitimate emails, the properties of 'illegitimacy' change in time as spammers strategically renew their contents to avoid recognition. In order to successfully identify and isolate such communications, spam classifiers need to be sensitized to the change in the textual properties of spam. Comparable concerns are found in customer behavior, weather forecast, and other phenomena where the predictive model loses effectiveness if not adaptively updated.

This paper addresses the challenge of classifying political texts in the presence of concept drift. In political science, automated textual analysis has been used to (1) identify keywords that characterize the frames of opposing parties on controversial policy issues [**?**]; (2) evaluate the nature of texts to gauge the policy position of the responsible parties [**?**]; . Since political actors compete over issue ownership and issue frame, how political groups identify themselves in political communication and how issues are presented are built on unstable, ever-changing images. As a result, many existing applications focus or rely on:

- events within a short time frame, such as parliamentary debates on a single bill;
- strong intuitions about the nature of conflict, i.e. issue-specific analysis;
- outputs that can be interpreted with reference to substantively specific contexts;
- fixed estimators, such as party position estimations without update.

The problem of concept drift for the classification of political texts is therefore uniquely linked to the phenomenon of agenda setting. In the policy process, attention flows from one policy issue to another [**?**] and policy issues move from one frame of reference to another over time [**?**]dery1984problem.

Traditionally viewed as an incremental process, agenda change appears to also experience abrupt shifts. Attention is allocated across topics disproportionately. Issue definition refers to a related phenomenon where actors use different symbols to craft issue frames. These symbols constitute the context in which imageries of policy issues are conveyed. The theoretical debates over the pattern of agenda change have not developed an operationalization of issue frames for quantitative micro-level analysis. In this respect, we consider lexical entities as a conduit of meanings in political discourse. A corpus of legislative deliberations in the French National Assembly is compiled. We examine the temporal changes in word distributions in this corpus using several measures such as occurrence probabilities and S scores. The results suggest that the temporal change in textual features is consistent with the broad descriptions of agenda change.

The challenge of studying agenda change partly arises from the data structure and from the research design. We identify several issues with the use of political texts in political science:

- Cases of agenda setting often focus on events within a short time frame, e.g. parliamentary debates over a single bill.
- The reliance on strong intuitions about the nature of conflict, i.e. issue-specific analysis.
- Textual analysis yields results that are only interpreted with reference to the political context, e.g. keywords by partisanship.
- The use of fixed frames to capture evolving concepts, e.g. party position measured on a right-left continuum.

2 Objectives and approach

Our task is to improve automated classification of political texts by using machine learning sensitive to concept drift. Using classifiers sensitized to concept drift will also enable the analytical evaluation of political texts as the product of a long-running and continuous process in which actors move from issue to issue and issues from frame to frame. Currently, we approach text classification in two forms of agenda change: the transition of actor attention from one issue to another over time (issue ownership) and the transition of frames (issue framing).

- **Issue ownership refers to the changing political association of policy issues.**
    1. Policy actors come to own a policy issue by successfully promoting their preferred issue frame against alternative imageries.
    2. For frequently-occurring terms, we consider their usage by actors affiliated with different organizational entities, i.e. political parties.
    3. We analyze the pattern in which selected frequent terms change in partisan affiliation. Inferring from the documented cases of agenda change, we expect the following patterns of change in issue ownership: (i) persistently owned by (associated with) only one political group; (ii) incrementally becoming less associated with one group and more associated with another group, and (iii) abruptly disowned or adopted by a group.
- **Issue framing refers to the characterization of policy issues using specific terms and symbols.**
    1. Political discourse is divided into issue-specific corpora using categories generated and assigned by the participants.
    2. Terms frequently occurring within each of these corpora are identified.
    3. The dynamic of issue framing is estimated as the year-to-year change in the membership of

3 Review of concept drift and its application in political science

4 Data collection and pro-processing

We compiled a corpus of legislative deliberations in the French National Assembly. [1]We examine word distributions in this corpus using S scores, *tf*, and *tf.idf* derived from the probabilities of word *w* occurring in each concept class. The results suggest that the temporal change in textual features is consistent with the broad descriptions of agenda change.

4.1 corpus

The core dataset is a diachronic corpus of verbatim records of chamber deliberations in the National Assembly of France between 2003 and 2010. Apart from the speeches delivered in the legislative chamber, the following information is extracted from this corpus:

- date of session
- document themes tabled for debate
- the identity of speakers
- the order of speeches

We stemmed the words and removed punctuation. The

For application I, the corpus is divided by partisan coalition (see table 1). UMP is the center-right coalition and SOC is the socialists and their allies.

Year	2003	2004	2005	2006	2007	2008	2009	2010
SOC	203	162	162	103	127	190	162	98
UMP	266	206	247	145	198	279	227	223
Total	469	368	409	148	325	469	389	321

Table 1: Corpus specifications: session count by political group

4.2 rubric-specific corpora

For application II, the corpus is divided into topic-specific sections using official categories, namely, finance act, social policy, and social security (see table 2).

Year	2000	2001	2002	2003	2004	2005	2006	2007
finance act	26	19	0	39	37	34	0	29
social policy	2	0	0	0	9	13	0	1
social security	9	9	0	6	8	6	0	5

Table 2: Corpus specifications: session count by topic-specific categories

The topics come from the official rubric nomenclature for the classification of legislative sessions by policy area. At the end of each legislative year, the archival services of the National Assembly publishes an annual report (*table analytique*), which contains a list of debates that took place in that year (*table chronologique des débats*).[2] We follow the assigned categorization and identify the above

[1]The corpus is developed from archives available at http://www.assemblee-nationale.fr/ and http://archives.assemblee-nationale.fr/

[2]The annual reports are publicly available at http://www.assemblee-nationale.fr/

sessions as dedicated entirely to the discourse on budgeting (*lois de finances*), social policy (*politique sociale*), and social security (*securité sociale*). A problem with this approach is that for sessions that cover more than one policy areas and those that deal with members' questions, the contents cannot be used for this application without further classification of content materials either automatically or manually. Also, many of the topics are not covered every year in legislative deliberations. Some issues only draw attention from lawmakers following major focal events, and that attention dissipates over time. Another problem is that some topics draw considerably more attention from the lawmakers than others. Depending on the clout of the policy constituents and their advocates, some issues can dominate the discourse. The asymmetry in attention allocation means that the majority of the rubrics are represented by very few sessions.

4.3 biographical data

The biographical and political information of the lawmakers is gathered from various official sources.

5 Temporal analysis results

In this section we present our results and show how word distribution changes across the time. We extract unigrams and bigrams and calculate their occurrence probabilities in each political coalition in the French parlimanent: the *Parti socialiste* (SOC) and the *Union pour un mouvement populaire* (UMP) and S scores.

Occurrence probabilities:

$$P_{soc}(w) = \frac{|d|w \in w|}{|P|}, where\ d \in P \tag{1}$$

$P_{soc}(w)$ is the set of documents d of the SOC class. $P_{soc}(w)$ computes the probability of w occurring in that set. $P_{ump}(w)$, for the occurrence in the UMP class, is similarly calculated.

S scores:

$$s(w) = |P_{soc}(w) - P_{ump}(w)|, where\ w \in w^D \tag{2}$$

and w^D is the set of all unique words in the corpus D. The score, $s(w)$, serves to measure the difference between $P_{soc}(w)$ and $P_{ump}(w)$ for the word w.

5.1 Application I: issue ownership

Issue ownership refers to the changing political association of policy issues. Political groups come to own a policy issue by identifying themselves with the issue in political discourse.

Inferring from the documented cases of agenda change, we expect the following patterns of change in issue ownership.

- Persistently owned by one political group;
- Incrementally changing affiliation, and;
- Abruptly disowned or adopted by a group.

Figure 1 shows temporal variation in s-cores of unigram between consecutive years of terms indicating varying levels of kurtosis: some years experience far more dramatic drifts. The variations may be a response to electoral events.

This initial analysis of word use in the French parliament supports the notion that idea changes over time and that evidence of such change could be accessed through textual analysis. The set of terms that come to define the input of a group in political discourse is unstable, such that many terms only

dominant the discourse for a limited time. The observed fluctuation is likely to be related to political processes such as electoral activities and policy events. For example, the terms union, popular, and mouv (which come from the name of the majority coalition, Union pour un mouvement populaire) occur at consistently high probabilities until after the general elections of 2008.

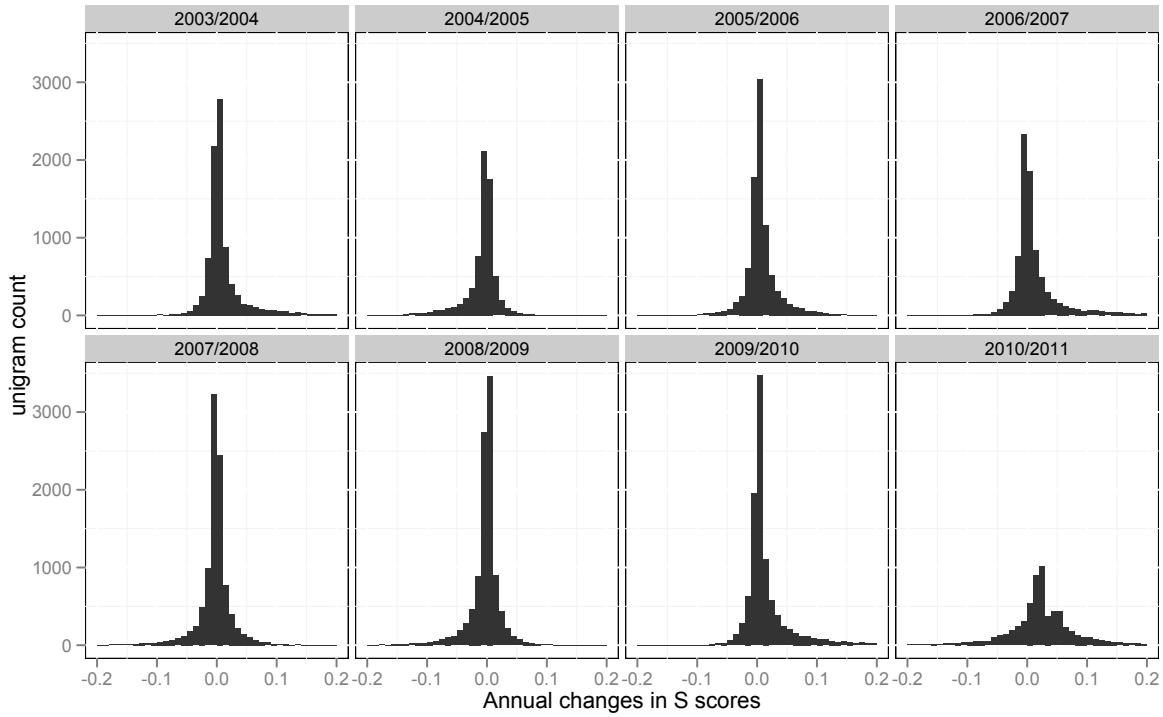

Figure 1: Year-to-year differences in the S scores of unique words indicating varying levels of kurtosis: some years experience far more dramatic drifts.

To find out about the drifting concept, we calculate the average value of the variation in probability score over the years. Table 3 contains 20 words that experience the highest average fluctuation in within-class probabilities.

In terms of fluctuations in within-class occurrence probability, we calculate the average difference of a term's year-specific within-class probabilities from its mean within-class probability and generate the list in table 3. Specifically,

$$\theta(w) = \frac{\Sigma(|P(w)_i - \overline{P(w)}|)}{years} \tag{3}$$

where $\theta(w)$ is the difference between the annual occurrence probabilities of word w in year i from the mean probability $\overline{P(w)}$ divided by the number of years in which w is observed in the corpus.

The lexical items listed in table 3 are the most unstable in their association with each of the two political groups in the French legislature and appear to be common terms in parliamentary debates, including names of political coalitions (e.g. 'popular', 'union', 'mouvement'), official titles (e.g. 'secretair', 'president'), years, and references to parliamentary procedures (e.g. 'major', 'system', 'assembl'). It is not immediately clear why these common terms experience high levels of variance in occurrence probability, but they may be linked to activities and processes characterized by strong seasonality (e.g. electoral activities, budgeting). The variations in common political terms and phrases are counterintuitive, and further analysis of the pattern is required to link the pattern to both regular and exceptional occurrences in political events and processes. In addition, the average extent of fluctuation is consistently larger for the UMP coalition than the socialists. This may be the result of corpus imbalance.

In terms of political salience, we identify several target words that have the greatest cross-class differences in occurrence probability. That is, for a given word w, its probability of appearing in a doc-

UMP	$\theta(w)$	SOC	$\theta(w)$
populair	0.3065416	coll	0.1973650
prem	0.2815950	union	0.1901781
diff	0.2547667	populair	0.1884494
anne	0.2465556	prem	0.1877775
coll	0.2450062	ration	0.1796245
ration	0.2410527	repr	0.1772250
general	0.2380369	rieur	0.1721089
rieur	0.2327067	ment	0.1703213
mouv	0.2293578	assembl	0.1694920
2009	0.2268894	sent	0.1675119
ment	0.2188756	mouv	0.1674238
secretair	0.2175922	collectiv	0.1593160
assembl	0.2153520	general	0.1591544
rent	0.2138963	diff	0.1590889
union	0.2128156	anne	0.1554741
particul	0.2118744	2004	0.1522438
2008	0.1962462	major	0.1508569
sident	0.1895400	consid	0.1485456
repr	0.1883056	tair	0.1397538
system	0.1865562	tabl	0.1355850

Table 3: Average score of variation in probability assignment by coalition

ument associated with political group i, represented as $S_i(w)$ in eq. 2, contrasts greatly with $S_j(w)$, its probability of appearing in a document associated with the (opponent) group j. Following the theoretical debates on policy agendas and advocacy, the extent of the difference may be indicative of the level of partisanship in word deployment - if members of one party invoke a certain term far more frequently than their parliamentary opponents, the term may represent an idea, philosophy, or perspective to which the party uniquely subscribes. Alternatively, the party in frequent use of a particular word may try to establish their ownership on the issue represented by that word. Another possible explanation is that the word represents a marginalized issue that is being actively promoted by a party. Where term preferences are strongly related to the users' political affiliation, these terms are likely to point to policy, ideological, and strategic differences that set those political groups apart. More importantly, our analysis shows that the set of these politically salient lexical entities is not stable across time: some terms are only associated with one party only for a limited time period, while others are persistently linked to one political group. Their temporal variation in salience may serve as an indicator of changing policy position or strategic arrangements among the political actors.

However, little about inter-party differences can be inferred from individual word stems. Each is generated in unspecified context, as a result of which the sense-in-use cannot be ascertained. Hence, further analysis of the concordance of these stems will be performed to obtain the necessary information. Considering their bigrams and word pairs may yield information about context. For example:

We also plot the the variation in party association probability and s-score of top 20 unigram specified in table 3. Figure shows how words changes it score among two parties across the year.

term	in English	mean salience (s score)	count
cadr	framework	0.4046753	8
afin	so as to	0.3920906	8
permettr	permit	0.3896127	8
objet	object (thing)	0.3702395	8
avis	view	0.3630753	8
egal	equal	0.3617324	7
permet	permit	0.3600077	8
appliqu	applied	0.3583102	8
object	object (goal)	0.3565936	8
ensembl	together	0.3556495	8
grac	thanks	0.3547573	7
adopt	adopt	0.3531716	8
renforc	reinforce	0.3517774	8
assur	assure	0.3492595	8
plac	place	0.3465509	8
titr	title	0.3454653	8
amelior	improve	0.3449567	7
adapt	adapt	0.3399255	8
tien	hold	0.3398084	8
environ	environment	0.3394885	8

Table 4: 20 words with the largest average s scores

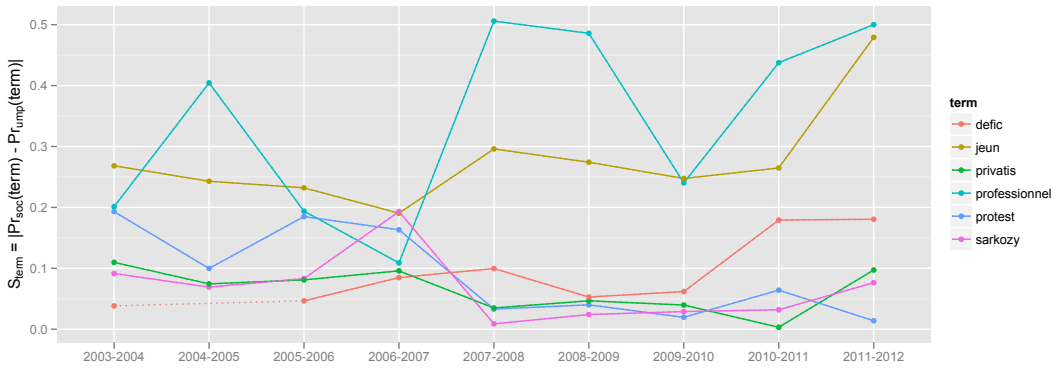

Figure 2: Terms that exhibit interesting patterns of variation in their association with the political groups over time. Some patterns are relatively incremental than others. Also, terms such as deficit and privatize appear more often in the speeches by the SOC members while youth and professional are more strongly linked to the UMP.

Figure 3: Top 20 words with the highest levels of fluctuation in within-class probability: UMP

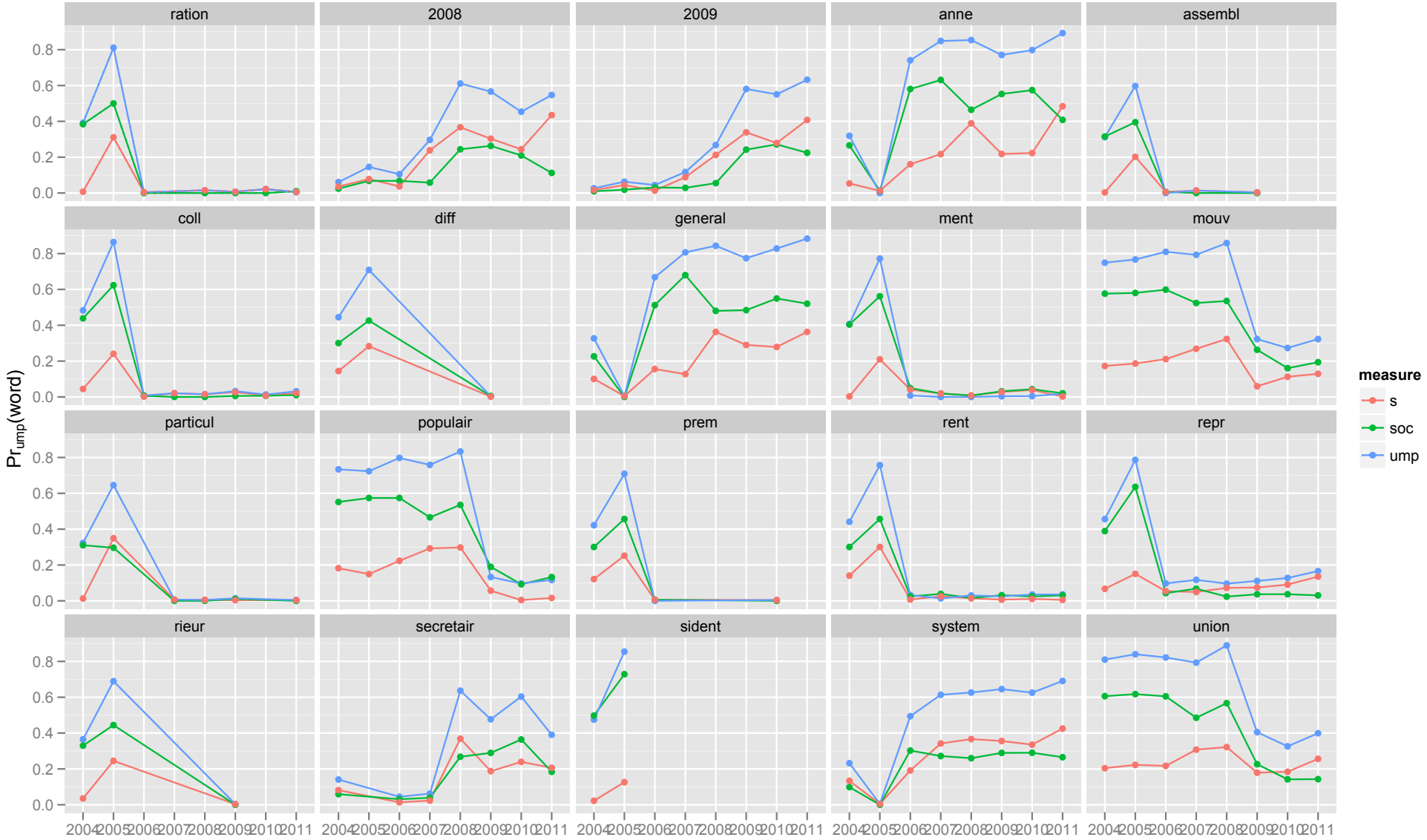

Figure 4: Top 20 words with the highest levels of fluctuation in within-class probability: SOC

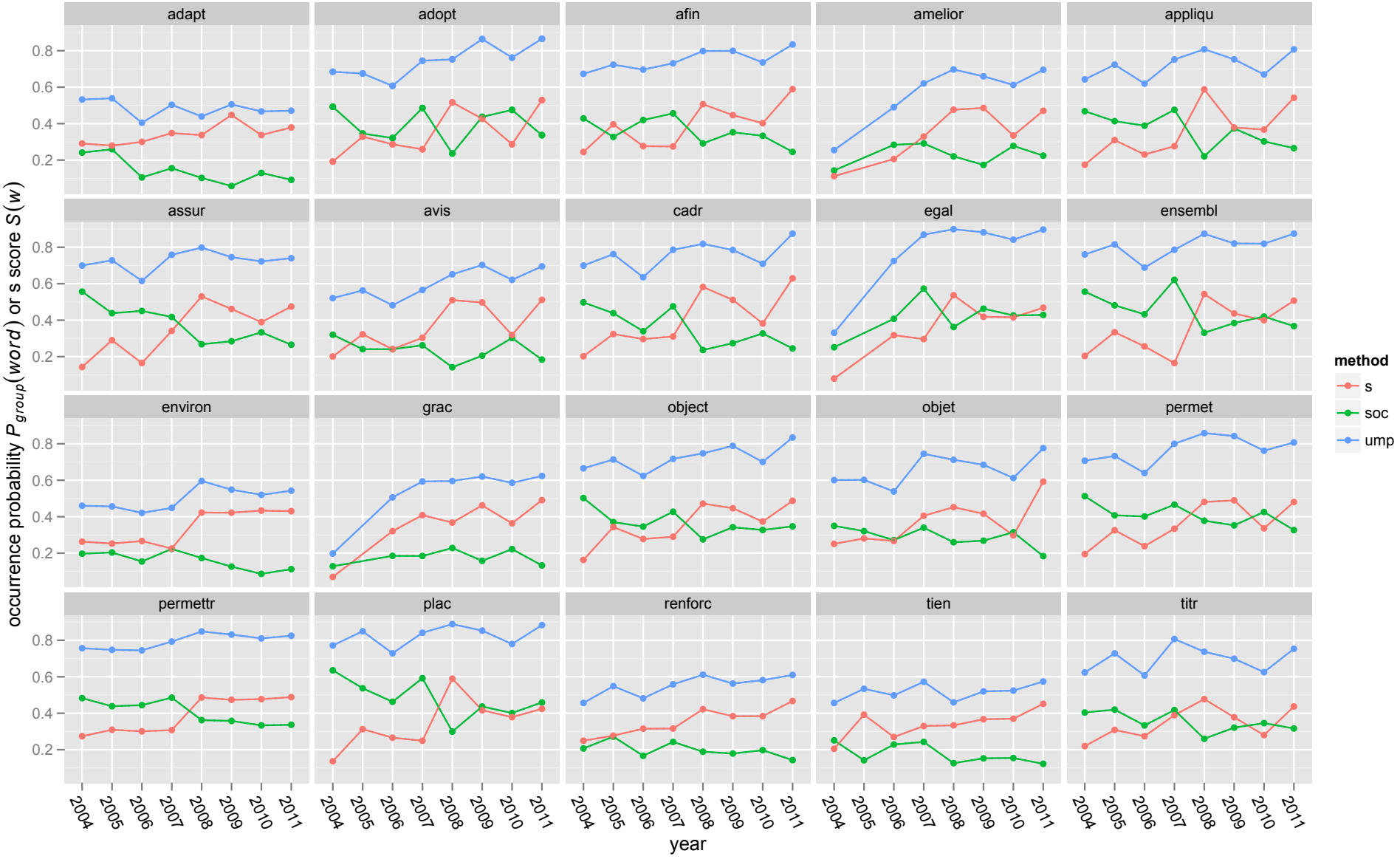

Figure 5: Top 20 words with the highest s scores

5.2 Application II: issue frame

Issue framing refers to the characterization of policy issues using specific terms and symbols. We develop issue-specific corpora using categories generated and assigned by the discourse participants. The S scores measuring occurrence within each of these corpora are identified and compared.

IN THE PLOT SAY ANOTHER WORD

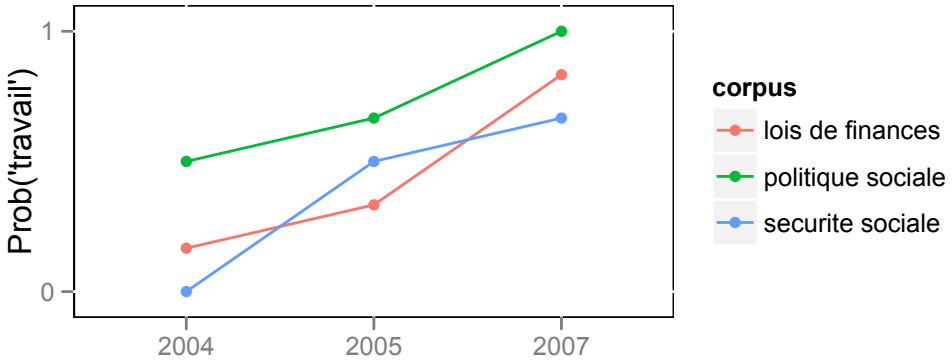

Figure 6: Figure 3. The word labor: changing association between the corpora of social policy, social security, and finance act. The term is increasingly invoked by actors in framing legislations in these areas.

Figure 3. The word 'labor': changing association between the corpora of social policy, social security, and finance act. The term is increasingly invoked by actors in framing legislations in these areas.

6 Classification results

7 Conclusion and future work

We need to enable a closer inspection and interpretation of the results by performing further analyses

1. Single words may not fully represent issues concordance. Also, the interpretation of unigrams is generally hard without any information on the context in which they are used. Some of the terms with the highest fluctuation in occurrence probability are expected to be relatively common. For example,

2. The use of top-down issue categories prevents the analysis of frames as entities independent of discourse events.

3. Connections between partisanship and frame to be specified.

4. Imbalance in the volume of speech across political groups.

5. Capture aggregate patterns of change emerging from those at the local (word) level.

6. Other political data may be brought into the analysis to substantiate claims that the drifts may be correlated with important national political and policy processes such as electoral activities, budgetary revisions, and media attention shifts. These data are available from on-going projects on policy agendas in the U.S. and France.

References

[1] F.R. Baumgartner and B.D. Jones. *Agendas and instability in American politics*. University of Chicago Press, 2009.

[2] D.A. Rochefort and R.W. Cobb. *The politics of problem definition: Shaping the policy agenda*. Univ Pr of Kansas, 1994.