SIFT – A Language Technology Toolkit to Assess the Print Media Coverage of New Forms of Governance

Bruno Wüest (Department of Political Science, University of Zurich)*
Michael Amsler (Institute of Computational Linguistics, University of Zurich)
Gerold Schneider (Institute of Computational Linguistics, University of Zurich)

November 2017
SIFT – A Language Technology Toolkit to Assess the Print Media Coverage of New Forms of Governance

Bruno Wüest¹, Michael Amsler² & Gerold Schneider²

¹ Department of Political Science, University of Zurich
² Institute of Computational Linguistics, University of Zurich

Abstract

For social scientists, it is increasingly important to explore large text collections without time-consuming human intervention. We are presenting a language technology tool kit that allows researchers of the NCCR Democracy Module 1 to extract information on various forms of governance from a comprehensive multilingual corpus. The tool kit called SIFT¹ allows searching for governance entities and measuring their salience, tonality, issue context and media frames. In substantial terms, our tool pipeline enables scholars of governance to extend their research focus to the previously neglected area of the public communication of democratic legitimacy and accountability of various forms of governance.

1 Introduction

Automated approaches to analyze unstructured text data have made tremendous progress in computational linguistics in the last decades (Jurafsky and Martin, 2009). They allow us to conduct research at a larger scale, with advanced statistical models and giving thorough quantitative support for intuition-based hypotheses. At the same time, social scientists are increasingly in need of such approaches, since the number of large, digitally available text collections is constantly growing. The research agenda in Module 1 of the NCCR Democracy (NCCRdm1) is no exception to this. Centered on researching the democratic legitimacy of new forms of governance, the NCCRdm1 focuses on a multitude of governance actors, policy fields and geographic areas. The amount of available and potentially relevant digital texts for this research agenda is tremendous, which diminishes the feasibility of manual content analysis procedures for large scale studies. The obvious task in this case is to transfer and adapt the comprehensive computational linguistic tool set to social science (Wüest et al., 2011). Accordingly, we present in the following a pipeline of language technologies that allows the analysis of big text data for research on the various forms of democratic governance.

The denationalization and privatization of democratic governance poses formidable challenges to the traditional, territorially grounded forms of democratic authorities (Zürn, 1998). At the European and international level, new modes of governance such as supra-national and intergovernmental bodies as well as transgovernmental networks have come to supplement classic intergovernmental governance (Abbott and Snidal, 2008). At the sub-national level, regulatory agencies and public-private partnerships increasingly spread across metropolitan regions by transforming traditional regional and local state institutions (Kelleher and Lowery, 2009).

These various new forms of governance have in common that they organize political authority along functional rather than territorial lines, which also implies that they are comparatively decoupled from representative democratic control. This is why some observers often declare see challenges for the accountability of single institutions or even more general for the democratic legitimacy for the political system as a whole (Follesdal and Hix, 2006; Keohane et al., 2009). Other scholars point to formal accountability mechanisms such as governmental and parliamentary over-sight as well as judicial review and highlight that they can at least partly compensate a deficit in democratic legitimacy (Lodge, 2002).

It is without controversy that the media are the most important source of information for most people in established democracies around...
the world (Bennett and Entman, 2001; Strömbäck, 2008; Walgrave et al., 2008; Müller, 2014; Arnold, 2004). But, politics has not only become predominantly mediated, the search for news value by the media also increasingly dominate political processes. This mediatization means that media attention increasingly influences the timing and substance of policy making processes.

Given this importance of the media, it is surprising that extant research in political science only rarely focuses on the role of the media for two core concepts of today’s established democracies: democratic legitimacy of the political system as a whole and the public accountability of single actors and institutions (but see Christmann et al., 2015; Maggetti, 2012 and Hurrelmann et al., 2009, for exceptions). The few studies that exist limit their focus to a few media outlets – most frequently, one or two outlets from the quality press are analysed – or entities – the existing studies are mainly case studies of a few actors or institutions. However, we are now in a situation where the advent of automated content methods enable the systematic analysis of large-scale text collections (Grimmer and Stewart, 2013).

The media can broadly hold two distinct functions for both democratic legitimacy and public accountability. On the one hand, the media can foster the quality of the different forums that debate and decide on legitimacy and accountability (Christmann et al., 2015; Jacobs and Schillemans, 2016). Hence, by conveying the necessary information, media provide the prime connecting mechanism between the different entities monitoring government and governance such as parliaments, courts and, of course, the citizenry and the actors and institutions that are to be held accountable. Empirical studies and theoretical accounts alike acknowledge the critical role of of freely operating media in maintaining a public sphere in which the performance of a democratic system and its actors is debated and negotiated (Habermas 2008, Hurrelmann).

On the other hand, the media has traditionally been conceived as an accountability forum in its own right, too (Graber, 2003). The media is more than a ‘watchdog’ in this perspective, since there are direct consequences if media uncover a violation of norms or standards (Christmann et al., 2015). Reputation is an indispensable resource for political actors in established democracies (Carpenter, 2010), so the limelight in the media can sometimes be enough to provoke corrective measures such as resignations of the responsible policy makers or reversion of policies.

There are, however, several limitations to both the supporting and the self-supporting function of the media for democratic legitimacy and public accountability. Most importantly, the media can never completely substitute formal processes of democratic control (Bovens, 2007b). Obviously, journalists and publishers are not elected. And they also do not possess any direct means to sanction or remove the rascals among the political elite. As welcome it is that media do regularly pay critical attention to governance processes, they can only encourage the formation of an informed public opinion (O’Donnell, 1998). It is then only the threat of electoral punishment or popular protest that mounts the ultimate pressure on governance actors to explain, justify and – if necessary – correct their conduct.

Further, the strive of the media to hold the powerful accountable may increase the risk of dysfunctional outcomes. With growing media attention, politicians are, for instance, inclined to care more about blame-avoidance strategies and a positive self-presentation than about genuine accountability (Papadopoulos, 2010).

In the following, we present SWIFT, a comprehensive corpus and a largely automated language technology pipeline, which enable political scientists to assess these questions. The paper begins by presenting our operationalization of indicators that allow the reliable measurement of governance accountability in a large-scale text analysis. Subsequently, we will describe the software pipeline and language technologies necessary to implement the operationalization, before we present a case study highlighting the feasibility of our approach.

2 Measuring media coverage on public accountability and democratic legitimacy

So far, mediatized accountability mechanisms have only been dealt with in conceptual elaborations or comparative case studies that entailed manual content analyses (Maggetti, 2012; Coglianese and Howard, 1998; Gerhards and Roose, 2007). Although these contributions are theoretically insightful and empirically rich, their focus on a narrow set of actors, geographical units
or media sources always faces the necessity to justify why their cases provide more than just idiosyncratic evidence. An automated large-scale analysis, by contrast, helps to achieve a more broad analytical support on the question whether and how media scrutinize on the accountability of governance processes.

2.1 Sample

The anchor of the analysis is a large gazetteer of pre-defined entities related to governance (see Figure 1). These entities refer to actors (collective actors and individuals), policy fields and regulation such as treaties or directives. At the moment, a comprehensive gazetteer of entities for 3257 queries is integrated in the document retrieval. The entities cover a large variety of forms of governance: transgovernmental networks, independent as well as private regulatory authorities, metropolitan bodies, supranational parliaments and international environmental governance outcomes. A gazetteer of entities needs to be hand-crafted for every new form of governance that is to be studied2.

<table>
<thead>
<tr>
<th>Project</th>
<th>Description of entities</th>
<th>Queries</th>
</tr>
</thead>
<tbody>
<tr>
<td>IP1</td>
<td>Metropolitan bodies on public transport and (local) economic promotion</td>
<td>26</td>
</tr>
<tr>
<td>IP2</td>
<td>International organizations, supranational parliaments and general international regulatory activity (directives, regulations etc.)</td>
<td>140</td>
</tr>
<tr>
<td>IP3</td>
<td>Regulatory activity on international environmental governance (treaties, conventions, protocols etc.)</td>
<td>216</td>
</tr>
<tr>
<td>IP4</td>
<td>Transnational networks (committees, institutions, associations, persons etc.)</td>
<td>238</td>
</tr>
<tr>
<td>IP5</td>
<td>Independent and private regulatory authorities (organisations, authorities, firms etc.)</td>
<td>57</td>
</tr>
</tbody>
</table>

Table 1: Entities of interest by research project

Most of the queries are additionally translated into the three languages (English, French and German), for which the analysis is set up. The European Economic Area, which is a relevant entity for the project IP2, for example, is searched as Espace économique européen in French sources and Europäischer Wirtschaftsraum in German sources. Also, where applicable, the queries have been extended with more general keywords such as abbreviations to a wide query. In English newspaper articles, the European Economic Area often is only mentioned with its abbreviation, EEA, which is why this is included into the search query as well.

The gazetteer of queries that map to governance entities is the only manual input of SIFT that is indispensable. After the governance entities of interests have been operationalized into search queries, the basic data generation therefore runs fully automated. Salience, Tonality and Issues can be generated automatically as well, whereas researches can optionally provide manual input for the topic models that calculate the issues. Only for the frame analysis, an active learning framework is implemented, which means that manual input in the form of labeled documents is required.

Figure 1: Stylized workflow in the language processing pipeline

In a first step, a comprehensive corpus of the following newspapers, newswires and online sources is established by retrieving all articles for the keyword gazetteer via API accesses to media content databases such as Lexis Nexis (1 in Figure 1).

- **Quality:** Frankfurter Allgemeine, Süddeutsche Zeitung, Welt, Tageszeitung (Germany); Figaro, Le Monde (France); Neue Zürcher Zeitung, Le Temps (Switzerland); The Guardian, London Times, Independent (UK)
- **Tabloid/Freesheets:** Bild (Germany); Aujourd’hui en France, 20 minutes (France); Blick, Le Matin, 20 Minuten (Switzerland); Daily Mail, Daily Mirror, Metro (UK)
- **Magazines:** Spiegel, Stern, Zeit (Germany); Nouvel Observateur, L’Express (France); Weltwoche, Wochenzeitung, L’Hedbo (Switzerland); New Statesman, Spectator, Economist (UK)
- **Regional:** Berliner Zeitung, Stuttgarter Zeitung, Stuttgarter Nachrichten (Germany); Le Parisien, Le Progrès (France); Tagesanzeiger, Berner Zeitung (Switzerland); London Evening Standard, City A.M., Birmingham Mail, Birmingham Post (UK)
- **Online sources:** Spiegel Online (Germany), Figaro Online, Le Monde Online (France); 20 Minuten Online (Switzerland); BBC News Online (UK)

---

2We are especially thankful to Michael Buess, Reto Wüst, Jofre Rocabet, Christian Ewert, Michael Strebel and Steffen Mohreberg who established, refined and delivered these gazetteers.
Since different types of media systems (Hallin and Mancini, 2004), as well as different types of media (Strömbäck and Kaid, 2008) possibly cover governance in different ways, the media sources are sampled so that there is a balanced set of outlets in our four countries (Switzerland, Germany, France, and United Kingdom). From each type of outlets, the outlet with the highest circulation (or website visits in the case of the online sources) was chosen. As far as possible, we also cover other potential variations such as different ideological leanings. In addition to these country-specific media samples, we also include a range of internationally operating newswires, which provide us with information on the general reporting on governance in disregard of specific journalistic cultures in single media outlets.

Subsequently, an additional layer of data consisting of the compressed documents along with initial meta-data (source, date-of-publication etc.) is added to the database (2 in Figure 1). At a third stage, we employ a full natural language processing chain, which includes morphological analysis, tagging, lemmatizing, and dependency parsing (3). Finally, a fourth layer of enhanced linguistic analysis – named entity recognition, coreference resolution, sentiment detection, opinion mining and topic modeling – is implemented to calculate the indicators of interest we will discuss in the following (4).

2.2 Salience
The attention media pay to specific forms of governance is the obvious starting point of the data analysis. No media attention is the worst case in terms of questions regarding the public accountability and legitimization of governance, since ‘quiet politics’ (Culpepper, 2010) implies absent interest by the public and, correspondingly, high leverage for particular interests and dishonest conduct in governance processes. The first necessary measure in the SWIFT analysis therefore is salience, defined as the visibility of specific governance entities in the media.

2.3 Tonality
A second crucial information on governance entities is the media’s evaluation of these governance entities in terms of tonality. The tone of media reports on governance entities yields useful results if changes in tonality signify reactions to events on the governance processes under concern (Maggetti, 2012). For example, if a corruption scandal shakes a governance actor, we expect media reports to shift to a negative tone. This also implies that tonality has to be measured at the level of the specific entity and not at the level of text documents as a whole.

We expect that the salience, tonality and issues in media reports on governance entities already reveal crucial evidence on whether and how media coverage entails mechanisms of accountability. More precisely, if media adjust their attention according to events related to specific governance entities, if media react to failure with a negative tone – and to success with a positive tone –, and if the media really cover the issues related to the area of responsibility of these governance entities, media coverage actually constitute an accountability forum’ for this governance entity (Bovens, 2007a).

2.4 Issues
Governance entities may draw media attention for different reasons, but not all are relevant for the research objective. If a sports magazine reported on the passion of the head of the Swiss Financial Markets Supervisory Agency (Finma) for windsurfing (which arguably is true), hardly any political analyst would deem this information relevant to understand financial market regulation in Switzerland. More generally, evidence on the thematic context in which governance entities are mentioned is key to assess whether media reports on specific entities are actually covering the governance processes of interest. In addition, an issue analysis caters to the growing interest in the study of the relationship between media and politics from an agenda-setting perspective (Vliegenthart et al., 2016).

2.5 Frames
What is still missing is information on the reasons why the media report on governance entities, i.e. which interpretations and problem definitions journalists convey to the reader. To this aim, we additionally conduct an analysis of the frames in the media (Entman et al., 2009). To start from the classical definitions, frames are schemata of interpretation (Goffman 1974, 21), or, very similarly, interpretive packages (Gamson...
and Modigliani 1989, 2) that are constructed in order to influence how people think about 'reality'. In the context of mega, we further specify frames as generic schemata of interpretation that refer to the main source of democratic legitimacy of governance entities as it is reported in the text documents. We are aware that many statements may not refer to any source of democratic legitimacy at all, however, for the study of media accountability of governance, evidence on the democratic legitimacy of governance is of the utmost significance.

To be precise, we separate input-oriented legitimacy frames from output-oriented ones. The concepts of output- and input-oriented legitimacy have their origin in Abraham Lincoln’s famous dictum about democracy requiring government by the people, of the people and for the people (Gopnik, 2007). Hence, for democratic governance to work properly, the participation of the potentially affected citizens, the representation of the citizen’s interests, and the authorities’ effectiveness is required. The labels for these concepts, in contrast, have diffused from systems theories, particularly from the work of Easton (1965), into the broader literature. Accordingly, the citizens’ support, be it directly through elections or indirectly through shared identities or diffuse system legitimization, constitute the input into the political system (see Schmidt 2013). Input legitimacy is thus present if media refer to participatory aspects, popular support and democratic accountability in general, or public interest representation with regards to governance processes. Output, on the other hand, is defined as the governments’ decisions and actions (Easton 1965). Output legitimacy, accordingly, refers to the efficiency, effectiveness and the due process of governance.

3 Empirical implementation and case study on the media coverage of the Kyoto protocol

3.1 Salience

For this case study, we measure salience as the occurrence of articles in the media coverage across the timeline. Although it is a simple measurement, salience reveals on the one hand important insights about the presence of the respective entity and, on the other hand, offers the opportunity to closer scrutinize the content near the peaks.

As can be seen in Figure 2, the visibility manifests itself with two clear peaks in 2007 and 2009. A closer investigation of the respective coverage points towards the importance of the Fourth Assessment Report of the United Nations Intergovernmental Panel on Climate Change (IPCC) in 2007 and the 2009 United Nations Climate Change Conference in Copenhagen which triggered each an increased attendance to the subject. We also see that most coverage stems from newswires, while print media are more selective in their reporting. Also, people who only have access to regional newspapers, we only observe a modest increase for the reporting of the main events.

In addition to this straightforward operationalization of salience, the number of articles found can additionally be weighted by document length. As an alternative to the number of documents, salience can also be expressed with the number of keyword hits, although we only recommend this for clearly delimitable entities.

3.2 Tonality

To measure tonality in the media coverage, we apply a linguistically informed sentiment analysis system, similar to (Taboada et al., 2011). More precisely, we use domain-adapted and extended lexical semantic resources in combination with a dependency parse that infers the syntactic structure of the sentence. The latter allows for composing tonality from singular words into chunks of the sentence (e.g., noun phrases) which in turn are then taken into account to derive the tonality analysis for the core of the sentences, i.e., the main verbs. A more detailed description can be found in (Klenner et al., 2014). The system used for this task was evaluated in another case study for the tracking of coverage tonality which yielded good results (see Wueest et al. (2014)). Although the tonality can be derived for singular entities in the
given texts, we aggregate in this case study on the document level since the thematical focus is narrowed by the data acquisition process (i.e. the query to the media databases).

Figure 3: Comparison of negative and ambivalent tonality between media types quality and newswire

In Figure 3 we focus on the difference of tonality regarding the level of critique considering different media types: the bars show the percentage of articles of negative and ambivalent tonality (ordinate on the left-hand side). It is obvious that the coverage in quality papers is much more critical than in the newswire articles. The lines show the percentage of only the ambivalent articles (ordinate on the right-hand side) which reveals that the difference between the two media types mainly stems from the much higher percentage of ambivalent articles, that is, articles which discuss the topic under different perspectives, considering chances and risks as well as progress and failure in the implementation process.

3.3 Issues

3.3.1 Topic Models

We apply structural topic models (STM) (Roberts et al., forthcoming) to explore the thematic context in which the media writes about governance. STM is a data-driven technique, which allows us to estimate document probabilities for latent variables, called topics. STM builds on the Latent Dirichlet Allocation, a hierarchical mixed-membership model in which the document-topic and word-topic probabilities have a common prior drawn from a Dirichlet distribution (Blei et al., 2003). One of the major innovations of STM is that the prior distribution of topics (i.e. topic prevalence) can be influenced by covariates. In the following analysis, we use the newspaper names and a b-spline with 10 degrees of freedom on a monthly trend variable to control for unwanted linguistic differences across news outlets and over time. In addition, we apply a parametric evaluation of the most probable topic-word vectors in order to find the optimal number of topics. To this purpose, we use word2vec (Mikolov and Dean, 2013), which learns and aggregates term similarities through a shallow neural network process. These term similarities can then be used to compare topic coherence and exclusiveness across different topic models. For the Kyoto protocol corpus, word2vec suggests a granularity of 19 for a candidate range of 3 to 20 topics.

Figure 4: Dynamics of selected topics

Figure 4 shows the trends in the prevalence of two especially meaningful topics over time. In addition, the list of the 10 most probable word stems for each topic is listed.

The first topic summarizes the different negotiations rounds on the Kyoto protocol, most notably the first commitment period from 2008 until 2012 with the Copenhagen summit in 2009 as key event. Reports on the different negotiations accordingly peak in this period. The second topic, in contrast, highlights the consequences of the Kyoto protocol on the energy markets and emission trading. Quite intuitively, this topic becomes most prevalent in the aftermath the big policy decisions from 2011 on.

3.3.2 Collocations

Issues can also be detected by various keyword algorithms. While they are less suitable to detect broad topics than topic modeling approaches, they reliably report keywords and key phrases. They typically perform better than topic modeling in situations where the amount of data is relatively small.

As a keyword detection method which does not need reference corpora, collocations measures can
be used to detect multi-word terms. Technical terms, also in less technical genres such as newspapers, are often multi-word concepts. According to Master (2003, 2), the multi-word units consisting of nouns are particularly important, as they are directly linked to neologisms and new concepts: “A noun compound is a grammatical structure in which nouns are linked together to indicate a new concept”. In information retrieval, multi-word expressions often offer a better unit of analysis than single words (Schwartz and Ungar, 2015, 84).

There exist a variety of collocation measures, see Evert (2009) and Pecina (2009) for a comprehensive overview. We use information-theoretic collocation measures based on observed frequency (O) of word pairs and expected frequency (E) of word pairs if the words in a text were randomly shuffled, i.e. the drawing of each of the two words in a collocation pair were an independent event. The corresponding measures, for example O divided by E (known as O/E or O over E) are easy to interpret.

Applied to the articles on the Kyoto protocol, O/E delivers the ranking given in Table 2. We only list the top 50 entries here, but the first several hundred entries contain many true positives, which can be easily filtered by manually sifting the lists. O/E has a well-known tendency to over-represent rare events (see e.g. Smith (2002)), which can be counteracted by using measures such as O^2/E (which scored best in Bartsch and Evert (2014)), O^3/E or measures based on significance testing, such as T-score. The top collocations for O^3/E are given as an example in Table 3. These lists can also be automatically filtered for named entities (such as place names or persons) or other key concepts related to the entity under concern.

### 3.4 Frames

While we have focused on purely empirical data for the other indicators, we will report on the evaluation for the methodological approach used for the framing prediction.

#### 3.4.1 Annotation

In contrast to the measurements for the other indicators which are derived generically, we rely on annotated data for the framing. More precisely, we annotate the frames using the brat annotation tool (Stenetorp et al., 2012). For the whole corpus three annotators build valuable training data of 18’548 frames\(^1\). After an intensive training phase, inter-annotator agreement is constantly high (micro-averaged F1-scores for fine-grained frame categories that range between 0.66

---

\(^{1}\)We thank Michelle Amman, Anna Sigrist and Anna-Lina Müller for their excellent work for the manual annotation data.
for 23 documents and during 0.71 for 5 documents at the start of the annotation).

3.4.2 Detection

First attempts have revealed that the recognition of frames is a challenging task, especially since we encounter a skewed distribution in the data (i.e. paragraphs containing frames vs. paragraphs without frames). Additionally, the distribution between the different types of frames is skewed as well (i.e. some frames occur much more than others), which then again complicates the task for a supervised learning approach.

Hence we design the automated recognition of frames as follows: in the first stage we apply a supervised learning approach. Secondly, we differentiate then between in-put frames output frames and throughput frames (i.e. paragraphs containing frames vs. paragraphs without frames). Lastly, we push the classifier in the ensemble scenario towards recall at cost of precision (system “rb”). In this way, we increase the recall of the frames up to 0.76.

Table 4 shows precision, recall, and F1 scores for the individual categories.

<table>
<thead>
<tr>
<th>Collocation</th>
<th>Rank</th>
<th>$O^3/E$</th>
<th>Collocation</th>
<th>Rank</th>
<th>$O^3/E$</th>
</tr>
</thead>
<tbody>
<tr>
<td>climate change</td>
<td>1</td>
<td>8.66e + 10</td>
<td>united states</td>
<td>26</td>
<td>1.38e + 09</td>
</tr>
<tr>
<td>kyoto protocol</td>
<td>2</td>
<td>2.29e + 10</td>
<td>de boer</td>
<td>27</td>
<td>1.34e + 09</td>
</tr>
<tr>
<td>per cent</td>
<td>3</td>
<td>2.05e + 10</td>
<td>long term</td>
<td>28</td>
<td>1.25e + 09</td>
</tr>
<tr>
<td>global warming</td>
<td>4</td>
<td>1.33e + 10</td>
<td>tony blair</td>
<td>29</td>
<td>1.17e + 09</td>
</tr>
<tr>
<td>greenhouse gas</td>
<td>5</td>
<td>1.04e + 10</td>
<td>the us</td>
<td>30</td>
<td>1.13e + 09</td>
</tr>
<tr>
<td>prime minister</td>
<td>6</td>
<td>9.08e + 09</td>
<td>to be</td>
<td>31</td>
<td>1.09e + 09</td>
</tr>
<tr>
<td>of the</td>
<td>7</td>
<td>7.16e + 09</td>
<td>it is</td>
<td>32</td>
<td>1.04e + 09</td>
</tr>
<tr>
<td>legally binding</td>
<td>8</td>
<td>5.51e + 09</td>
<td>barack obama</td>
<td>33</td>
<td>1.02e + 09</td>
</tr>
<tr>
<td>don t</td>
<td>9</td>
<td>4.29e + 09</td>
<td>george bush</td>
<td>34</td>
<td>9.79e + 08</td>
</tr>
<tr>
<td>greenhouse gases</td>
<td>10</td>
<td>4.19e + 09</td>
<td>saudi arabia</td>
<td>35</td>
<td>9.73e + 08</td>
</tr>
<tr>
<td>developing countries</td>
<td>11</td>
<td>3.36e + 09</td>
<td>rather than</td>
<td>36</td>
<td>9.67e + 08</td>
</tr>
<tr>
<td>carbon dioxide</td>
<td>12</td>
<td>3.00e + 09</td>
<td>power stations</td>
<td>37</td>
<td>9.58e + 08</td>
</tr>
<tr>
<td>fossil fuels</td>
<td>13</td>
<td>2.96e + 09</td>
<td>nni michael mccarthy</td>
<td>38</td>
<td>9.50e + 08</td>
</tr>
<tr>
<td>in the</td>
<td>14</td>
<td>2.86e + 09</td>
<td>nr blair</td>
<td>39</td>
<td>8.82e + 08</td>
</tr>
<tr>
<td>such as</td>
<td>15</td>
<td>2.41e + 09</td>
<td>have been</td>
<td>40</td>
<td>8.66e + 08</td>
</tr>
<tr>
<td>intergovernmental panel</td>
<td>16</td>
<td>2.25e + 09</td>
<td>world s</td>
<td>41</td>
<td>7.83e + 08</td>
</tr>
<tr>
<td>will be</td>
<td>17</td>
<td>2.24e + 09</td>
<td>framework convention</td>
<td>42</td>
<td>7.75e + 08</td>
</tr>
<tr>
<td>has been</td>
<td>18</td>
<td>1.97e + 09</td>
<td>european union</td>
<td>43</td>
<td>7.55e + 08</td>
</tr>
<tr>
<td>the world</td>
<td>19</td>
<td>1.90e + 09</td>
<td>downing street</td>
<td>44</td>
<td>7.44e + 08</td>
</tr>
<tr>
<td>ki moon</td>
<td>20</td>
<td>1.89e + 09</td>
<td>ndavid adam</td>
<td>45</td>
<td>7.26e + 08</td>
</tr>
<tr>
<td>tar sands</td>
<td>21</td>
<td>1.84e + 09</td>
<td>trading scheme</td>
<td>46</td>
<td>7.08e + 08</td>
</tr>
<tr>
<td>gordon brown</td>
<td>22</td>
<td>1.77e + 09</td>
<td>the kyoto</td>
<td>47</td>
<td>7.03e + 08</td>
</tr>
<tr>
<td>white house</td>
<td>23</td>
<td>1.75e + 09</td>
<td>al gore</td>
<td>48</td>
<td>6.91e + 08</td>
</tr>
<tr>
<td>more than</td>
<td>24</td>
<td>1.68e + 09</td>
<td>fossil fuel</td>
<td>49</td>
<td>6.80e + 08</td>
</tr>
<tr>
<td>gas emissions</td>
<td>25</td>
<td>1.50e + 09</td>
<td>angela merkel</td>
<td>50</td>
<td>6.67e + 08</td>
</tr>
</tbody>
</table>

For the frame detection in paragraphs in the first stage, we apply the following preprocessing steps. First, we embed the sentences of the paragraphs using word embeddings derived from the full text.
corpus of the project for all three languages. Note that we filter here for content words only as there is evidence that this enhances the process of sentence content representation (cf. (Lau and Baldwin, 2016)). Second, we compare the embedded sentence to quantized representations (i.e., clustered to 10 centroids per frame category) of the aforementioned concept features as well as we calculate raw counts and ratios for those features. Based on this comparisons, we produce scorings for each frame category on the fine-grained level as well as on the coarse-grained level. After the preprocessing step, we train an soft voting ensemble classifier, consisting of a random forest classifier, a logistic regression, and a nave bayes estimator. To furthermore counteract the data imbalance leading to a low recall for the frames, we down-sample the paragraphs not containing a frame according to their distribution in the feature space (see “rb” version in Table 4).

It must also be noted that the baseline approach which is bag-of-word based is far less transferable to other projects since it makes heavy use of domain adaptation via the used vocabulary. However, since the requirements of the involved projects are different, we allow for shifting the decision boundary of the classifier ex-post to reach a higher overall accuracy while maintaining an acceptable recall for the frames. Table 5 shows an example for such a threshold based scenario where we reach an overall accuracy above 72% while keeping the recall for frames above 0.5.

<table>
<thead>
<tr>
<th></th>
<th>Prec.</th>
<th>Rec.</th>
<th>F1</th>
<th>Prec.</th>
<th>Rec.</th>
<th>F1</th>
<th>Acc.</th>
<th>TH</th>
</tr>
</thead>
<tbody>
<tr>
<td>EN: rbT</td>
<td>0.26</td>
<td>0.61</td>
<td>0.36</td>
<td>0.93</td>
<td>0.75</td>
<td>0.83</td>
<td>0.73</td>
<td>0.7</td>
</tr>
<tr>
<td>FR: rbT</td>
<td>0.34</td>
<td>0.53</td>
<td>0.42</td>
<td>0.88</td>
<td>0.76</td>
<td>0.82</td>
<td>0.72</td>
<td>0.8</td>
</tr>
<tr>
<td>DE: rbT</td>
<td>0.26</td>
<td>0.54</td>
<td>0.35</td>
<td>0.91</td>
<td>0.75</td>
<td>0.82</td>
<td>0.72</td>
<td>0.8</td>
</tr>
</tbody>
</table>

Table 5: Evaluation for the detection of frames using a threshold.

3.4.3 Classifying Frames

While the system predicts frames on the unit of a paragraph, we identified during the annotation the slice of the text which was the core of the frame, i.e., the text passage that triggered the annotation of the presence of the specific legitimacy frame subclass. Since this piece of information contains the most secure information of how such a frame is textually manifest, we rely on this spans to create the seeds for the aforementioned lexicon generation process. To corroborate the robustness of the system to differentiate on the coarse-grained and fine-grained level we evaluate on this textual passages only. Table 6 shows that the system performs above the accuracy of 0.8 on coarse-grained level and between 0.58 and 0.62 on the fine-grained level. Note that we report weighted average values.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>EN</td>
<td>0.81</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
<td>0.79</td>
<td>0.59</td>
<td>0.58</td>
<td>0.58</td>
<td>0.58</td>
<td>0.58</td>
<td>0.58</td>
<td>0.58</td>
<td>0.58</td>
<td>0.58</td>
<td>0.58</td>
<td>0.58</td>
<td>0.58</td>
<td>0.58</td>
<td>0.58</td>
</tr>
<tr>
<td>FR</td>
<td>0.83</td>
<td>0.83</td>
<td>0.83</td>
<td>0.83</td>
<td>0.61</td>
<td>0.62</td>
<td>0.61</td>
<td>0.62</td>
<td>0.62</td>
<td>0.62</td>
<td>0.62</td>
<td>0.62</td>
<td>0.62</td>
<td>0.62</td>
<td>0.62</td>
<td>0.62</td>
<td>0.62</td>
<td>0.62</td>
<td>0.62</td>
<td>0.62</td>
</tr>
<tr>
<td>DE</td>
<td>0.83</td>
<td>0.83</td>
<td>0.83</td>
<td>0.83</td>
<td>0.60</td>
<td>0.61</td>
<td>0.60</td>
<td>0.60</td>
<td>0.60</td>
<td>0.60</td>
<td>0.60</td>
<td>0.60</td>
<td>0.60</td>
<td>0.60</td>
<td>0.60</td>
<td>0.60</td>
<td>0.60</td>
<td>0.60</td>
<td>0.60</td>
<td>0.60</td>
</tr>
</tbody>
</table>

In contrast, the general prediction task is carried out on the unit of the paragraph, not the frame text passage itself. Nevertheless, the identification of the latter is crucial to derive the resources on which the system relies.

3.4.4 Generalization and Transferability

To test the transferability of the framing detection and classification component, we evaluated the performance separately on another annotated dataset which was sampled from entities of government which were not included in the dataset we developed and trained on. This dataset contains another 1’516 paragraphs which were annotated on the fine-grained level. We report here an overall accuracy for the frame detection of 0.72. If we predict directly the coarse-grained level for paragraphs, we score at 0.63. For the fine-grained level, the accuracy is 0.54.

4 Conclusion

The corpus and language technology pipeline SIFT was built around the assumption that the salience, tonality and issues in media reports on governance entities reveal crucial evidence on whether and how media support or diminish democratic legitimacy and public accountability of governance. In this paper, we discussed the methodological decisions and the technical implementation necessary to acquire and analyse about 600 governance entities in roughly 4.5 mio. news articles. The case study on the Kyoto protocol has shown, that we can draw meaningful conclusions from the results provided by SIFT. At least under certain conditions, media adjust their attention according to events related to specific governance

---

4 We thank Rolf Badat for his great work for the manual annotation as well the evaluation of the analysis.
entities; react to failure with a negative tone – and to success with a positive tone –; and convey the issues and frames related to the area of responsibility of the respective governance entity.

We have shown how news documents can be mined to uncover the media’s role for public accountability and democratic legitimacy. More substantially oriented research is necessary to make full use of SIFT in order to further disentangle this role. Possible research questions, to name only a few, are whether and to what effect media coverage varies across different governance entities; whether increased monitoring by the media has the intended, positive effect on legitimacy and accountability; or whether media are merely an aggregator and amplifier for other legitimacy and accountability forums or constitute an own, independent forum.

Acknowledgments

Research for this paper was funded by the NCCR Democracy at the University of Zurich.

References


