University of Zurich[UZH]

selects
MEDIENANALYSE

Department of Political Science

Selects Media Analyses 2015

Election Campaign in Swiss National Media

Codebook & Technical Report

Bruno Wüest, Sarah Bütikofer, Adrian van der Lek, Fionn Gantenbein

Table of Contents

Swiss Election Studies – Media Analyses Election Campaign

Swiss Election Studies – Media Analyses Election Campaign

1 Background and Aim of the project

1.1 Swiss Election Studies Selects

The Swiss Electoral Studies Selects is a research project conducted by FORS, the Swiss Foundation for Research in Social Sciences, and funded by the Swiss National Science Foundation. The aim of the Selects project is to study the voting behavior in Switzerland in depth.
In connection with the Rolling Cross-Section Analyses conducted by Selects, it was decided to integrate an additional media analyses of the Swiss National Election Campaign 2015.

1.2 Selects Media Analyses 2015

The media analyses follows two goals. First, it allows investigating the content of the election campaign in the Swiss media preceding the 2015 federal elections. Second, it makes it possible to refine the analysis of the opinion formation process during the various phases of the campaign.
In contrast to previous media content analyses, the 2015 Selects Media Analysis relied on an automated content analysis method. This allowed broadening the scope of the analysis, by covering a much larger number of newspaper articles and by including more sources, from all Swiss regions.

The Selects Media analyes selected all articles related to politics and published during the election campaign, and coded them by identifying the most important actors (politicians and parties) and topics.

1.3 Selects Media Analyses 2015 Data Sets

The present codebook offers information on the different data sets that were constructed, the additional variables that were integrated, and a description of the data sets themselves. In addition, a technical report describes the methods used and the various steps of the Media Analyses.

2 Data

2.1 Overview

The Selects Media Analyses 2015 integrated a total of 45'863 articles from 93 different Swiss Newspapers and Magazines[1], published between 1st August August (Swiss national holiday, start of the election campaign) and 18 October 2015 (National election day).
The goal of the media analyses was to identify the most important actors (persons and parties) and topics during the electoral campaign.

2.2 Datasets

Table 1 presents an overview of the eight different datasets. Each one is designed for a specific type of analysis. The datasets can be matched to each other using common variables such as the document identifier 'docid', the party keyword identifier ('party_n') or the person identifiers ('person_n').

Table 1: Overview of all datasets

Name of Dataset	Available Data Format	Content	Observations
Selects-Media-Overview	.csv, .dta, .rds	This dataset contains the basic information of all 45'863 media documents integrated into our analyses.	See chapter 2.2.1 for variable description and chapter 4.1 for a complete list of all media sources
Selects-Media-Parties	csv, .dta, .rds.	This datasets contains the number of hits for all party keywords in the respective media articles.	See chapter 2.2.2 for variable description and 4.2 for the list of all party keywords.
idx_parties.csv	csv	This dataset contains the list of all party keywords and the respective category.	This is an additional dataset that contains information at the level of every single party. It may be matched to other datasets.
Selects-Media-Persons	csv, .dta, .rds.	This dataset contains the information of the number of hits for the 3'927 persons (politicians) mentioned in the electoral campaign.	Match idx.persons for more information on the persons included into the analyses.
idx_persons.csv	csv	This dataset contains additional information on the persons (politicians)	This is an additional dataset that contains information at the level of every single person. It may be matched to other datasets.
Selects-Media-Topics	csv, .dta, .rds.	This dataset contains the information on the probability of every identified topic.	See 4.3 for a complete overview of all topics
Selects-Media-Polls	csv, .dta, .rds.	This dataset contains the information on the poll keywords	
idx_polls.csv	csv	This dataset contains additional information of the poll keywords	

[1] Please find the list of all integrated media titles in chapter 4.1

2.2.1 Overview media analysis dataset

Table 2: Selects-Media-Overview

Variable name	Description
docid	Document identifier, 8 digits
source	Abbreviated name of source (official short name used by the Swiss Media Database SMD)
source_txt	Name of source (Media title, see list of all media in chapter 4.2)
publdate	Date of publication of the media article
lang	Language of the publication
length	Length of document in number of words
title	Title of the media article
totaldocs	Number of articles published by the same source on the same day

2.2.2 Datasets for parties

Table 3: Selects-Media-Parties

Variable name	Description
docid	Document identifier, 8 digits
source	Abbreviated name of source (official short name used by the Swiss Media Database SMD)
publdate	Date of publication of the media article
lang	Language of the publication
party_1	Number of hits for party keyword with identification "party_1"
party_2… party_226	number of hits by party keyword (total 226 party keywords, see list of all parties in chapter 4.2 and the additional data at the level of parties "idx_parties.csv")
party_al	Aggregated number of party keyword hits for "Alternative Linke"
party_others	Aggregated number of party keyword hits for "others" (compare to list of all parties in chapter 4.2).
party_bdp	Aggregated number of party keyword hits for "BDP"
party_cvp	Aggregated number of party keyword hits for "CVP"
party_edu	Aggregated number of party keyword hits for "EDU"
party_evp	Aggregated number of party keyword hits for "EVP"
party_fdp	Aggregated number of party keyword hits for "FDP"
party_glp	Aggregated number of party keyword hits for "GLP"
party_gps	Aggregated number of party keyword hits for "GPS"
party_lega	Aggregated number of party keyword hits for "Lega"
party_mcg	Aggregated number of party keyword hits for "MCG"
party_rl	Aggregated number of party keyword hits for "small left parties" (compare to list of all parties in chapter 4.2).
party_rr	Aggregated number of party keyword hits for "small right parties" (compare to list of all parties in chapter 4.2).
party_sp	Aggregated number of party keyword hits for "SP"
party_svp	Aggregated number of party keyword hits for "SVP"

Table 4: idx_parties.csv

Indicator name	Description
idx	Identification of party (reference to data set, especially the single variables on party keywords, see "party_...")
key	Regular Expression as used in the analyses
label	Name of party keyword
party	Abbreviation of party (see the single variables on parties, i.e. "party_AL" ..., in the data set). Categories: CVP, GPS, OTHERS, RL, LEGA, SP, FDP, RR, EDU, AL, BDP, EVP, GLP, MCG, SVP.

2.2.3 Datasets for persons

Table 5: Selects-Media-Persons

Variable name	**Description**
docid	Document identifier, 8 digits
source	Abbreviated name of source (official short name used by the Swiss Media Database SMD)
publdate	Date of publication of the media article
lang	Language of the publication
person_1	number of hits for first person keyword (Name of the politician)
person_2… person_3927	number of hits for second, etc. person keyword (total 3'927 persons, see detailed information of all persons the additional data at the level of parties "idx_persons.csv")
person_glp_runningIncumbent … person_al_candidate	Aggregated number of person keyword hits for the following parties: CVP, GPS, OTHERS, RL, LEGA, SP, FDP, RR, EDU, AL, BDP, EVP, GLP, MCG, SVP and the following functions: federal councilor, running incumbents, resigning incumbents and candidates

Table 6: idx_persons.csv

Indicator name	Description
idx	ID of party keyword (reference to data set, especially the single variables on party keywords, see "party_...")
hits	Total of hits of the person within the election campaign in all media titles integrated in the analyses
lastname	Last name of the person
firstname	First name of the person
partyshort	Party short name of party person belongs to
gender	1 if female, 0 if male
canton	Canton (constituency)
parliamentID	Identification of MP, assigned through Swiss Parliamentary Services
selectsID	Identification number in Selects Data Set
elected in National Council	1 if person is elected in National Council, 0 if not
elected in Gouvernment	1 if person is elected in Federal Council, 0 if not
elected in Council of States	1 if person is elected in Council of States, 0 if not
deselected	1 if person is deselected, 0 if not
federalCouncillor	1 if person is a Federal Councillor, 0 if not.
runningIncumbent	1 if person is a running incumbent, 0 if not
resigningIncumbent	1 if person is **not running** during federal election campaign 2015, 0 if

	not
candidate	1 if person is a running for office for the first time

2.2.4 Datasets for topics

See chapter 4.3 for a complete list of all topics.

Table 7: Selects-Media-Parties

Variable name	Description
docid	Document identifier, 8 digits
source	Abbreviated name of source (official short name used by the Swiss Media Database SMD)
publdate	Date of publication of the media article
lang	Language of the publication
de_topic1	Probability that the article can be assigned to the topic with identification "de_topic1" (first topic for German speaking media titles)
de_topic2 … de_topic17	Probability that the article can be assigned to the topic with identification from "de_topic2" to "de_topic17"
fr_topic1	Probability that the article can be assigned to the topic with identification "fr_topic1" (first topic for French speaking media titles)
fr_topic2 … fr_topic18	Probability that the article can be assigned to the topic with respective identification (from "fr_topic2" to "fr_topic18")
it_topic1	Probability that the article can be assigned to the topic with identification "it_topic1" (first topic for Italian speaking media titles)
it_topic2 … it_topic18	Probability that the article can be assigned to the topic with respective identification (from "it_topic2" to "it_topic18")

2.2.5 Datasets for polls

Table 8: Selects-Media-Polls

Variable name	Description
docid	Document identifier, 8 digits
source	Abbreviated name of source (official short name used by the Swiss Media Database SMD)
publdate	Date of publication of the media article
lang	Language of the publication
poll_233	number of hits for the first poll keyword
poll_234 … poll_250	number of hits for the second poll keyword to the last poll keyword
poll	aggregated number of hits for all poll keywords

Table 9: idx_polls.csv

Indicator name	Description
idx	ID of poll keyword (from poll_233 to poll_250)
key	Regular Expression as used in the analyses
label	Name of poll keyword

3 Technical Report

3.1 Short description of the automated media content analyses

The main tasks for the Selects 2015 Media Analysis comprise the download, preprocessing and storage of large amounts of media documents, the identification of relevant documents (classification), the extraction of party, person and poll keywords (named entity recognition) and the generation of semantic classes (topic models). The pipeline consists of two separate KNIME[2] workflows that are run in succession, as well as three batches of external scripts written in Python or R.

KNIME-based stages

Download. Fetches, preprocesses and filters documents from the archives of the Swiss Media Service (SMD). The user specifies the language, time-span and a list of publication codes to be either included or excluded from the raw dump. **Also,** minor clean-up and data splitting tasks are conducted at this point.

Classification and named entity recognition. Applies a previously trained scikit-learn-based ensemble classifier and a dictionary-based named entity recognition script to the data.

External stages and supplementary scripts

Sampling. A separate workflow was used to sample a training set from the article corpus, using a random stratified sampling on the corpus.

Classifier training. Testing and training of the ensemble classifier is done using four external Python scripts.

Topic models. Finally, the Structural Topic Model is run using an external R script, which takes the output data of the second KNIME workflow as its input.

3.2 Work steps of the analyses

3.2.1 Download

The downloader takes a start and end date, a language shortcut, as well as a file specifying the publications to be included into the media corpus. It then performs the following steps:

1. In order to accommodate the download limit of 10'000 documents per request, a list of URLs covering spanning sufficiently small date intervals (as to stay below the cap) is dynamically generated.
2. The workflow subsequently loops over this list of URLs, fetching and preprocessing the specified article corpus into chunks.
3. Another embedded Python script authenticates with the HTTP-based Apache Solr/Lucene interface of the SMD using a virtual browser and downloads the respective chunk.
4. Following this step, the received XML is split into separate documents. A series of XPath queries retrieves values from the relevant fields (title, body, publication, publication date, etc.) and stores the data in KNIME's internal database format.

[2] KNIME, the Konstanz Information Miner, is an open source data analytics, reporting and integration platform, see https://www.knime.org/.

5. Finally, a filter excludes or includes publications according to a file provided by the user, before the data is written to disk in a KNIME-specific database dump format.

A variant of this workflow allows to fetch articles by their publication ID's.

In total, we retrieved 209'090 documents from German language sources, 63'424 documents from French language sources and 3'193 from Italian language sources.

3.2.2 Classification

The first stage of the analyses is concerned with the selection of relevant documents, which is done by binary classification. In an initial phase, three different definitions of relevant were considered: classifying documents into those relevant for politics in general, those relevant for Swiss politics only as well as those relevant for the federal election campaign as such. After a pre-test, we settled for the middle ground, which is classifying document according to their relevance for Swiss politics. While the general politics category clearly is too broad to be relevant to the Swiss federal elections, we also discarded a classification of documents for their relevance to the electoral campaign as such. In a pre-test, we found that it was already very difficult for human annotators to achieve a decent agreement on what exactly counts as relevant. Besides this, a practical reason against a selection by relevance for the federal election campaign is that the number of documents retrieved would drop to very low numbers for the Italian language documents, which would make further automated analyses impossible.

Training data

The first step is the compilation of a training corpus in each of the three languages. The samples consist of a random stratified draws on the corpus, where each stratum corresponds to a publication. This stratification ensures that we train the classification on the full variety of the language used by different media types. The total size of the corpus, in turn, corresponds to the distribution of downloaded documents, although the Italian language sources were up-weighted in order to have enough documents in this language as well. The following table reports on the key figures of the training data.

	German language documents	*French language documents*	*Italian language documents*
Total sample	1'813	978	395
Positives (relevant for Swiss politics)	432	146	114
Share of positives	23.8%	15.0%	28.8%

The following definition guides the manual annotation into Swiss politics:

"A document reports on Swiss politics if it is published as an editorial document, opinion or commentary on war and conflict, official matters affecting one or more states, elections and votes, adoption of laws, political issues, public reforms in various policy areas and other matters directly relating to politics.

If this condition is met, a document is considered to cover Swiss politics, as soon as a Swiss actor (official person or organization) occurs, Switzerland is mentioned or domestic Swiss policies are the subject of the document.

No coverage of Swiss politics are all other documents as well as documents, which could be classified as Swiss politics because of their content, but which do not report on strictly political matters. The latter means, for example, stories on an exhibition about the First World War, the review of a book on the political system of Switzerland, and the like."

In order to establish the applicability of this definition, we conducted an interannotator-agreement test among three expert coders (two postdoc researchers and one Ph.D. student) on a random sample of 100 documents. On average, a recall of 0.67, precision of 0.93 and an according f1 score of 0.78 was achieved. After this pre-test, we rewrote the definition of Swiss politics and retrained our coders in order to enhance the quality of the annotations especially in terms of recall.

Ensemble training

In order to generalize from the training data to the full corpus of our analyses, we build and train ensemble classifiers for each language. Four different classifiers from the Python scikit-learn machine learning library are thereby considered candidates for the final ensemble classifier: Kernel Ridge (KR), Stochastic Gradient Descent, using a linear Support Vector Machine as loss function (SGD), Multinomial Naive Bayes (MNB) and Random Forests (RF). All four classifiers were trained and tested for a general F1-score, which provides a balanced scoring over recall and precision. For the SGD, MNB and RF classifiers, it was additionally possible to include an additional version that optimizes for the recall of the positives in order to accommodate the skewed occurrence of positives (only about 10-20% of all documents). For each classifier, a grid of possible parametrizations has been defined. Using randomized search cross-validation, the best parameter set out of 20 bootstrapped iterations was determined and stored.

Each classifier is embedded in its own pipeline, consisting of the following stages:

Stage	*Description*	*Varying Parameters*
Count vectorizer	Preprocessing step. Converts a document into a token count matrix.	- lowercase all characters - stop word filter from the Python NLTK Toolkit - stemming tokenizer (removing word suffixes) - ignore terms with a frequency per document - limit the vocabulary size (including only the top n features ordered by term frequency) - use word n-grams
tf-idf-Transformer	Normalizes the token counts from the previous stages using the tf-idf measure (term frequency–inverse document frequency)	- enable or disable inverse document frequency - normalize term vectors or not
Classifier	The actual classifier	- Alpha (multiplier of the regularization term) - Gamma (influence of single training examples) - Class weights (expected share between positives and negatives in the data) - Penalty (regularization term to be used, either the squared Euclidean norm, the absolute norm or a combination of both (elastic net); SGD only) - Shuffle (shuffling of training data after each epoch of parameter optimization; SGD only) - L1 Ratio (elastic net mixing parameter; SGD only) - Criterion (split quality measure, gini or entropy; RF only) - Maximal depth of tree (RF only) - Bootstrapping (RF only)

Subsequently, the selected parameter sets were re-tested with 10-fold random sampling for four different splits of the training corpus into test and training sets, in order to verify stability of the classification results and to determine the most suitable classifiers for a classification ensemble.

Testing of the ensemble occurred using a more elaborate cross-validation method. The performance was evaluated across nine splits. First, ten subsets were randomly sampled from the training data. For each split, a fixed number of ten folds was generated. The folds consisted of all (non-overlapping) sliding windows across the subset, "padded" with additional combinations of subsets.

For German language documents, an ensemble of the F1-score optimized MNB and both the F1-score and the recall optimized SGD was chosen. For French language documents, a combination of F1-score optimized MNB, SGD and RF classifiers was most satisfactory. For Italian language documents, finally, F1-score optimized MNB, SGD and RF classifiers were used along with the recall optimized SGD and RF.

Finally, using the Python package 'dill', each model was serialized and written to disk, ready for usage in the third KNIME workflow.

Evaluation

We present an out-of-sample as well as a post-prediction evaluation of the performance of the three ensembles to classify the German, French and Italian language documents described in the last section. The following table lists recall, precision and F1-scores for an out-of-sample evaluation on a held-out set of documents from the training data.

	precision	*recall*	*F1-score*	*N*
German language documents				
Negatives	0.94	0.97	0.96	276
Positives	0.90	0.80	0.85	87
Weighted average	**0.93**	**0.93**	**0.93**	**363**
French language documents				
Negatives	0.95	0.94	0.95	167
Positives	0.68	0.72	0.70	29
Weighted average	**0.91**	**0.91**	**0.91**	**196**
Italian language documents				
Negatives	0.88	0.91	0.89	54
Positives	0.78	0.72	0.75	25
Weighted average	**0.85**	**0.85**	**0.85**	**79**

Most importantly, as the numbers for the weighted averages over negatives (i.e. not Swiss politics articles) and positives indicate, the classifiers overall perform satisfactory. It is also worth noting that their performance exceeds the manual inter-annotator pre-test, which can be perceived as a benchmark for this task.

On closer inspection, it gets clear that the ensembles perform substantially better at identifying negatives, i.e. irrelevant documents, for all languages. This is plausible since the ensembles had more manually classified documents to learn from for the negatives (their share is 76.2%, 85% and 71.2% in the training corpus). While the difficulty to identify positives is not severe for the German language documents, it seems a more demanding challenge for the French and Italian language documents. However, the amount of positives in the test data is also very low here (29 and 25 documents), so that some few ambivalent annotations might heavily bias the results. This is why we also conducted a post-predictive evaluation.

After the ensemble classifications were carried out, we randomly sampled and manually checked 100 positively classified documents for German and 50 for French and Italian. We are thus able to assess the precision on the actual results of the ensembles. It is 0.90 for German, 0.92 for French and 0.70 for Italian. Hence, the post-predictive evaluation yields about the same evidence for the German language documents than the out-of-sample evaluation, a much better result for French, and a worse but still acceptable result for Italian.

In sum, the share of negatives that are wrongly classified as positives (as indicated by the precision) as well as the amount of positives which are lost (as indicated by the recall) is even lower (for German and French) or at about the same level (for Italian) than in the manual pre-test. As for French and German language documents, nine of ten documents can be assumed to report on Swiss politics. As for Italian, still seven out of ten documents are expected to be relevant ones. Here, however, the low numbers in the overall corpus as well as the training data seems to make the task more difficult. After the application of the ensemble classifiers, the numbers are as presented in the following table.

	German language documents	*French language documents*	*Italian language documents*
N full corpus	209’090	63’424	3’193
N positives	38’468	7’438	348
Share of positives	18,4%	11,7%	10,9%

It is the positives, which are included into the subsequently introduced analyses. The different shares of positives compared to the manual annotations might at least partly due to the stratified sampling strategy in the manual data. The oversampling of some sources evidently led to an overestimation of the share of positives in the manual annotation.

3.2.3 Named entity recognition (NER)

Our NER task is to extract the number of mentions of Swiss parties, relevant Swiss politicians as well as mentions of election polls, i.e. the entities for the recognition, from the documents. We first attempted using the Stanford Named Entity Recognizer for German and French (using a Python client) and the NER component of the TextPro NLP suite for Italian. After modest preliminary results, it was determined that a more straightforward approach based on dictionary lookups and regular expressions was preferable. This strategy proved especially feasible since all entities of interest were readily available. Hence, we merged the official list of candidates in the federal election 2015 running for seats in the National Council or the Council of States with our own list of Federal Councillors, party presidents and resigning incumbents. The final dictionary features 3'913 politicians (see the separate documentation on persons). As for the party keywords, we started from the list names in all party-lists submitted by the parties for the federal election 2015. From this extensive list, we compiled regular expressions by removing duplicates, reducing the party names to their minimum (e.g. "Lega" instead of "Lega dei Ticinesi"), and by anticipating different cases (e.g. "Christichen Demokratischen Volkspartei" and "Christichen Demokratischen Volkspartei"). Subsequently, we added alternative descriptors to the list, which we established in research on Twitter data on Swiss politics (e.g. "Freisinn" for the "FDP.Die Liberalen"; see Wueest, Müller and Willi, 2016). In a last step, we extensively tested the list of regular expressions in an interactive way on the regular user interface of the SMD newspaper database. The resulting dictionary features 181 party specific keywords (see the separate documentation on parties). Finally, we assembled 18 keywords that signal reporting on electoral polls (names of polling institutes as well as general keywords on polls).

In the end, the NER proved to be task for which the setup of a comprehensive dictionary was possible. Since all entities were tested several times, we also expect a high precision of the keywords in the general results.

Regular expression-based NER

The regular-expression-based implementation is written in Python and makes use of Google's RE2 high-performance regular expression library to reduce run-time. The script is parallelized externally using parallelization nodes provided by the KNIME Labs plugin.

The script loads party keywords, as well as names of politicians from separate CSV files. Party names that contain characters with accents are additionally transliterated to an ASCII-compliant character set. For persons with multiple first names, nicknames or surnames, all possible combinations are stored, again with transliterated versions where necessary. Named Entity candidates are identified using a large, OR-ed, pre-compiled regular expression. Using a secondary look-up, the starting indices of all occurrences of these candidates are determined. In addition, matches are normalized, as the initial match was case-insensitive. The mentions are mapped to the canonical name (raw combination of first name and last name for persons, label for parties), which in turn is mapped to its corresponding ID.

3.2.4 Topic models

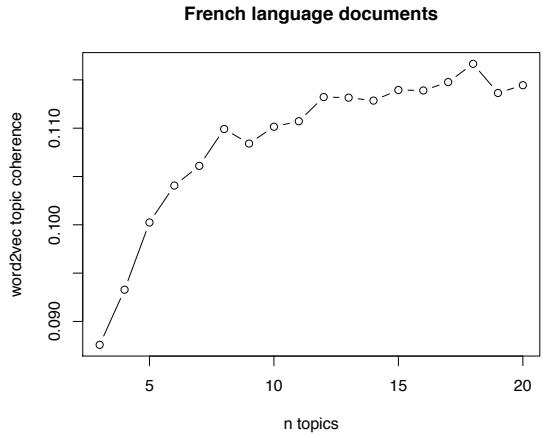
We identify the semantic structure of our corpus of relevant with a series of structural topic models (STM) (Roberts et al., 2014), which allows us to estimate document probabilities for latent semantic variables, called *topics*. STM builds on the Latent Dirichlet Allocation, a well-established hierarchical mixed-membership model in which the document-topic and word-topic probabilities have a common

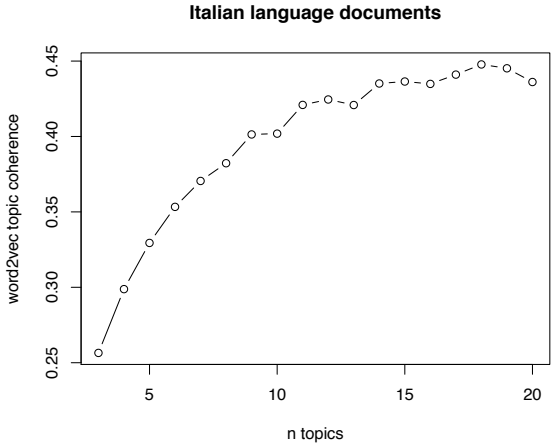
prior drawn from a Dirichlet distribution. Subsequently, every word in every document is drawn from a multinomial distribution, conditional on the word's probability of being drawn from the different topics.

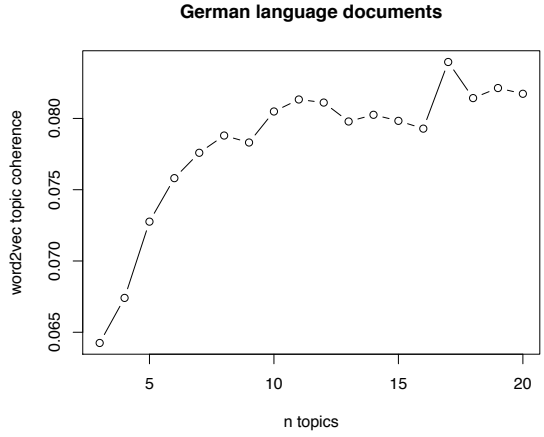
The fact that the LDA is a mixed-membership model means that it assumes that each document consists of a mixture of topics. Hence, probabilities are estimated for each document-topic combination. One of the STM's major innovations is that the prior distribution of topics can be influenced by covariates. In our analyses, we include the publication dates and source titles as covariates in order to control for unwanted influence of linguistic shifts over time and between different sources.

The STM model is fitted with a semi-collapsed variational Expectation-Maximization algorithm, which yields estimates of the quantities of interests, in our case the document-topic probabilities as well as the word-topic rankings. Prior to the estimation, we pre-processed all documents with standard procedures such as tokenizing, removal of punctuations and stop-words, as well as stemming and converting all words to lowercase.

A crucial decision in every application of a topic model pertains to the granularity, i.e. the number of topics. A topic model with too few topics will produce overly broad, diffuse topics, while a model with too many topics will results in many small, hardly distinguishable topics. We tackle this issue by comparing the coherence of the word rankings generated by different topic models. To this purpose, we use *word2vec* (Mikolov et al., 2013), which learns and aggregates term similarities through a shallow neural network process. By comparing the coherence within and between the vectors of most probable words for each topic model, word2vec suggests a granularity of 18 (French and Italian) and 17 (German) for a candidate range of 3 to 20 topics (see the following figures).







We use the word-topic rankings to identify the substance of the different topics. Our understanding of the substance as well as the 30 most probable words are indicated in the separate documentation on topics. In addition, we read some few high probable documents for every topic in order safeguard against misleading interpretations of the word ranking profiles.

4 References

Mikolov, T., K. Chen, G. Corrado, and J. Dean (2013). "Efficient Estimation of Word Representations in Vector Space," CoRR, abs/1301.3781.

Roberts, M. E., B. M. Stewart, D. Tingley, C. Lucas, J. Leder-Luis, S. Gadarian, B. Albertson and D. Rand (2014). "Structural topic models for open-ended survey responses." American Journal of Political Science 58: 1064-1082.

Wueest, B, C. Müller and T. Willi (2016). "Exploring the usefulness of Twitter data for political analysis in Switzerland", paper prepared for the Annual Conference of the Swiss Political Science Association at the University of Basel, January 21-22, 2016.

Wueest, B, S. Bütikofer, F. Gantenbein, A. van der Lek (2016). "Selects Medienanalyse 2015: Der Wahlkampf 2015 in den Schweizer Medien". Zurich: IPZ.

5 Appendix

5.1 List of the media sources included in the analyses

Indicator name	Description
newsShort	Abbreviation of the media source (reference to data set)
newsName	Name of the media source
newsCirculation	Newspaper's circulation (source WEMF Bulletin 2015)
newsType	Type of news source (labels: freesheet, news portal, supraregional, regional paper, business paper, local paper, business news portal, boulevard, weekly paper, company paper, tabloid magazine, sunday paper, online)
newsLangauge	Publication language of news source

newsShort	newsName	newsCirculation	newsType	newsLanguage
AGE	Agefi	NA	business paper	fr
AVU	Anzeiger von Uster	41757	local paper	de
AZM	Aargauer Zeitung	86458	regional paper	de
BAZ	Basler Zeitung	53498	regional paper	de
BIT	Bieler Tagblatt	21739	local paper	de
BIZO	Bilanz online	NA	business news portal	de
BLI	Blick	163627	boulevard	de
BLIA	Blick am Abend	284771	boulevard	de
BODU	Bote der Urschweiz	30559	local paper	de
BU	Bund	44411	regional paper	de
BUET	Bündner Tagblatt	8124	regional paper	de
BZ	Berner Zeitung	152974	regional paper	de
BZM	Basellandschaftliche Zeitung	135502	regional paper	de
CASO	Cash online	NA	business news portal	de
COOI	Cooperazione	123159	company paper	it

COOP	Coopzeitung	1818588	company paper	de
EXIM	Express	18431	regional paper	fr
FN	Freiburger Nachrichten	40912	regional paper	de
FURT	Furttaler	15379	local paper	de
FUW	Finanz und Wirtschaft	25067	business paper	de
FUWO	Finanz und Wirtschaft online	NA	business news portal	de
GHI	GHI	NA	freesheet	fr
GLAT	Glattaler	27075	local paper	de
GP	Glückspost	156098	tabloid magazine	de
HEB	Hebdo	38325	weekly paper	fr
HEU	24 heures	65505	supraregional	fr
ILLE	Illustré	80344	tabloid magazine	fr
INFS	Infosperber	NA	news portal	de
JJ	Journal du Jura	9364	regional paper	fr
JMO	Journal de Morges	39796	local paper	fr
LB	Landbote	84853	regional paper	de
LBH	Broye	24951	weekly paper	fr
LIB	Liberté	39828	regional paper	fr
LTZ	Limmattaler Zeitung	8176	local paper	de
MEWO	Medienwoche	NA	news portal	de
MM	Migros-Magazin	1569115	company paper	de
MME	Migros Magazine	506306	company paper	fr
NLZ	Neue Luzerner Zeitung	124355	supraregional	de
NLZS	Zentralschweiz am Sonntag	99513	sunday paper	de
NNBE	Newsnet / Berner Zeitung	NA	news portal	de
NNBS	Newsnet / Basler Zeitung	NA	news portal	de
NNBU	Newsnet / Der Bund	NA	news portal	de
NNHEU	Newsnet / 24 heures	NA	news portal	fr
NNTA	Newsnet / Tages-Anzeiger	NA	news portal	de
NNTDG	Newsnet / Tribune de Genève	NA	news portal	fr
NNTLM	Newsnet / Le Matin	NA	news portal	fr
NOU	Nouvelliste	39200	regional paper	fr
NZZ	Neue Zürcher Zeitung	124043	supraregional	de
NZZS	NZZ am Sonntag	135805	sunday paper	de
OAS	Ostschweiz am Sonntag	59005	sunday paper	de
OLT	Oltner Tagblatt	14594	local paper	de
ONA	Obersee Nachrichten	68822	local paper	de
RTS	Newsplattform	NA	news portal	fr
RUEM	Rümlanger	3696	local paper	de
SAS	Schweiz am Sonntag	199624	sunday paper	de
SBAU	Schweizer Bauer	30540	company paper	de
SBLI	Sonntagsblick	188302	sunday paper	de
SEBO	Seetaler Bote	5034	local paper	de
SF	Schweizer Familie	194427	tabloid magazine	de
SGT	St. Galler Tagblatt	128519	supraregional	de
SHZ	Handelszeitung	37909	business paper	de
SHZO	Handelszeitung online	NA	business news portal	de
SI	Schweizer Illustrierte	188197	tabloid magazine	de
SOS	Südostschweiz	81302	supraregional	de
SOZM	Solothurner Zeitung	22207	regional paper	de
SRF	srf.ch	NA	news portal	de
SWII	swissinfo.ch	NA	news portal	de
TA	Tages-Anzeiger	172920	supraregional	de
TAGZ	Tagblatt der Stadt Zürich	121566	local paper	de
TAM	Magazin	368376	weekly paper	de
TAS	SonntagsZeitung	201738	sunday paper	de
TAWO	TagesWoche Online	NA	news portal	de

TAWP	TagesWoche	23846	regional paper	de
TDG	Tribune de Genève	43860	supraregional	fr
TLM	Matin	47934	supraregional	fr
TLMD	Matin Dimanche	135609	sunday paper	fr
TPS	Temps	37021	supraregional	fr
TZ	Thurgauer Zeitung	34200	regional paper	de
VOLK	Volketswiler	27075	local paper	de
WB	Walliser Bote	32463	regional paper	de
WEOB	Werdenberger & Obertoggenburger	20407	local paper	de
WEW	Weltwoche	58430	weekly paper	de
WILB	Willisauer Bote	9333	local paper	de
WOZ	Wochenzeitung	15867	weekly paper	de
ZHOL	Zürcher Oberländer	97038	regional paper	de
ZHUL	Zürcher Unterländer	77162	regional paper	de
ZOF	Zofinger Tagblatt	12476	local paper	de
ZPLU	zentral+	NA	news portal	de
ZSZ	Zürichsee-Zeitung	31032	local paper	de
ZWA	20 minuten	476638	freesheet	de
ZWAI	20 minuti	34071	freesheet	it
ZWAO	20 minuten online	NA	news portal	de
ZWAS	20 minutes	199142	freesheet	fr

5.2 List of the party keywords

Indicator name	Description
Party ID	Party identification number
Party Name	Official name of the party in the respective canton / election list
Party Short Name	Aggregation to a party family, short name

partyID	Party Name	Party Short Name
party_1	Alternative Linke	AL
party_10	Partito Pirata	OTHERS
party_100	JFDP	FDP
party_101	jfs	FDP
party_102	JLR	FDP
party_103	PLR	FDP
party_104	JGLP	GLP
party_105	Grün-Liberal	GLP
party_107	Vert'Libéral	GLP
party_109	Verdi Liberali	GLP
party_11	Patriotisch Liberale Demokraten	OTHERS
party_111	GLP	GLP
party_112	PVL	GLP
party_114	Giovani Verdi	GPS
party_115	Grüne Panther	GPS
party_116	Grüne Partei	GPS
party_118	Jeunes Vert-e-s	GPS
party_119	Jeunes Verts	GPS
party_120	Junge Grüne	GPS

party_121	Grünes Bündnis	GPS
party_123	Les Seniors Verts	GPS
party_124	Les Verts	GPS
party_126	Ökoliberal	GPS
party_127	Partié Écologiste	GPS
party_128	Partito Ecologista	GPS
party_129	BastA!	GPS
party_13	Piraten Partei	OTHERS
party_130	GP	GPS
party_131	JGS	GPS
party_132	JVS	GPS
party_133	POP Verts Sol	GPS
party_134	Lega Dei Ticinesi	LEGA
party_136	Lega	LEGA
party_137	Mouvement Citoyens Romands	MCG
party_138	Mouvement Citoyens Genevois	MCG
party_139	MCG	MCG
party_14	Sozial-Liberale Bewegung	OTHERS
party_140	MCR	MCG
party_141	Alliance de Gauche	RL
party_142	Ensemble à Gauche	RL
party_143	Grüne Unabhängige	RL
party_144	Jeunesse Communiste	RL
party_145	Junge Alternative JA!	AL
party_146	Kommunistische Jugend	RL
party_147	Partei der Arbeit	RL
party_148	Parti Ouvrier et populaire	RL
party_149	Parti Suisse du Travail	RL
party_15	Anti-PowerPoint-Partei	OTHERS
party_150	Partito Comunista	RL
party_151	Partito Operaio e Popolare	RL
party_152	CS-POP	RL
party_153	PDA	RL
party_154	PNOS	RR
party_155	Action Nationale	RR
party_156	Aktive Senioren	RR
party_157	Démocrates Suisses	RR
party_158	Democratici Svizzeri	RR
party_159	Direktdemokratische Partei Schweiz	RR
party_16	Avenir et Réflexions	OTHERS
party_160	Freiheitspartei	RR
party_161	Impossible Alternative	RR
party_163	Mouvement Démocratique Cadmos	RR
party_164	Nationale Aktion	RR
party_165	Parti Nationaliste Suisse	RR
party_166	Schweizer Demokraten	RR
party_167	Suisse de la Liberté	RR
party_168	Svizzero della Liberta	RR
party_169	Volks-Aktion	RR

party_17	Die Unpolitischen	OTHERS
party_170	Alpenparlament	RR
party_171	DPS	RR
party_172	FPS	RR
party_174	PNS	RR
party_175	SD	RR
party_177	JUSO	SP
party_178	GISO	SP
party_179	Gioventù Socialista	SP
party_18	Ecopop	OTHERS
party_180	Jeunes Socialistes	SP
party_183	Jungsozialist	SP
party_184	Les Socialistes	SP
party_185	Parti Socialiste	SP
party_186	Partito Socialista	SP
party_187	Sozialdemokratich	SP
party_188	JUSOplus	SP
party_189	PS	SP
party_190	second@	SP
party_191	SP	SP
party_192	JSVP	SVP
party_193	JUDC	SVP
party_194	Schweizerischen Volkspartei	SVP
party_195	Union Démocratique du Centre	SVP
party_196	Unione Democratica di Centro	SVP
party_197	GUDC	SVP
party_198	JUSVP-CH	SVP
party_199	PDP	SVP
party_2	Integrale Politik	OTHERS
party_20	Graines de Futur	OTHERS
party_200	SVP	SVP
party_201	UDC	SVP
party_21	Indépendants Vaudois	OTHERS
party_22	IP	OTHERS
party_222	PES	GPS
party_223	Verts	GPS
party_224	Grüne	GPS
party_225	GPS	GPS
party_226	Verdi	GPS
party_23	Kunst+Politik	OTHERS
party_24	LDP	OTHERS
party_25	Liste du Vote Blanc	OTHERS
party_26	LOVB	OTHERS
party_27	LPS	OTHERS
party_29	mach-politik.ch	OTHERS
party_3	Jungliberale	OTHERS
party_30	Montagna-Viva	OTHERS
party_31	NPL	OTHERS
party_32	ÖBS	OTHERS

party_33	Ökoliberale Bewegung	OTHERS
party_34	Piratenpartei	OTHERS
party_35	Pirates	OTHERS
party_36	Politique Intégrale	OTHERS
party_37	Rauraque du Nord	OTHERS
party_38	SLB	OTHERS
party_39	Tierpartei Schweiz	OTHERS
party_4	Liberal-Demokratische Partei	OTHERS
party_41	Unabhängigkeitspartei up!	OTHERS
party_42	Zentrumspartei	OTHERS
party_43	Zukunft und Reflexion	OTHERS
party_44	Borghese-Democratico	BDP
party_45	Bourgeois-Démocratique	BDP
party_46	Buergerlich-Demokratisch	BDP
party_47	BDP	BDP
party_48	JBDP	BDP
party_49	PBD	BDP
party_5	Lösungs-Orientierte Volks-Bewegung	OTHERS
party_51	Chrétien-Social	CVP
party_52	Christdemokratisch	CVP
party_53	Christlich-Sozial	CVP
party_54	Christlich-demokratisch	CVP
party_56	Christlichsoziale Volkspartei	CVP
party_57	Communauté de travail économie et Société	CVP
party_58	Démocrate Chrétien	CVP
party_59	Democratico-Cristiano	CVP
party_6	Nouveau Parti Libéral	OTHERS
party_60	Generazione-Giovani	CVP
party_62	Jeunes PDC	CVP
party_63	Partito Cristiano Sociale	CVP
party_64	Popolare Democratico	CVP
party_65	Arbeitsgemeinschaft Wirtschaft und Gesellschaft	CVP
party_66	AWG	CVP
party_67	C-Partei	CVP
party_68	CSP	CVP
party_69	CVP	CVP
party_7	Parteifreie	OTHERS
party_70	GDC	CVP
party_72	JCSP	CVP
party_73	JCVP	CVP
party_74	JDC	CVP
party_77	PCS	CVP
party_78	PDC	CVP
party_79	PPD	CVP
party_8	Parteilose Schweizer	OTHERS
party_80	Eidgenössisch-Demokratische Union	EDU
party_81	Union Démocratique Fédérale	EDU
party_82	Unione Democratica Federale	EDU
party_83	EDU	EDU

party_84	JEDU	EDU
party_85	UDF	EDU
party_86	JEVP	EVP
party_87	Evangelische Volkspartei	EVP
party_88	Parti-Evangelique	EVP
party_89	Partito Evangelico	EVP
party_9	Parti Pirate	OTHERS
party_90	EVP	EVP
party_91	PEV	EVP
party_92	Freisinn	FDP
party_93	Jungfreisinn	FDP
party_94	Liberali Radicali	FDP
party_95	Liberaux Radicaux	FDP
party_96	Umweltfreisinnig	FDP
party_97	FDP	FDP
party_99	jf	FDP

5.3 List of all topics

Indicator name	**Description**
Variable Name	Name of the topic (topics are different for the three languages)
Topic Name	Name and description of the topic

Variable Name	**Topic Name**
de_topic1	Elections
de_topic2	Electoral Campaign
de_topic3	Energy policy
de_topic4	Local affairs
de_topic5	Education policy
de_topic6	Federalism
de_topic7	Church affairs (Bishop Huonder)
de_topic8	Law making process and direct democracy
de_topic9	Local affairs
de_topic10	Law making process
de_topic11	Transportation policy
de_topic12	Local affairs Berne
de_topic13	Quality of life
de_topic14	Strength of Swiss franc
de_topic15	Asylum policy and refugees
de_topic16	Rule of law
de_topic17	Bilateral relations Switzerland EU
fr_topic1	Bilateral relations Switzerland EU
fr_topic2	Elections
fr_topic3	Education policy

fr_topic4	Corporate tax reform
fr_topic5	Electoral Campaign
fr_topic6	Swiss People's Party Canton of Vaud
fr_topic7	Fiscal policy
fr_topic8	Institutional affairs and direct democracy
fr_topic9	Local affairs
fr_topic10	Energy policy
fr_topic11	Service public in Switzerland
fr_topic12	Politics and democracy
fr_topic13	Rule of law
fr_topic14	Geneva and Islam debate
fr_topic15	Health policy
fr_topic16	Strength of Swiss franc
fr_topic17	Asylum policy and refugees
fr_topic18	Labor market and pension policy
it_topic1	Assisted suicide
it_topic2	Middle East conflict
it_topic3	Local affairs in the Canton of Ticino
it_topic4	Burka ban
it_topic5	Asylum policy and refugees
it_topic6	Rassism in Social Media
it_topic7	Whistleblowing and bank secrecy
it_topic8	Direct democracy and elections
it_topic9	Fiscal policy and corruption
it_topic10	Youth policy
it_topic11	E-voting and direct democracy
it_topic12	Elections and electoral campaign
it_topic13	Consumerism
it_topic14	Economic policy
it_topic15	Fiscal policy
it_topic16	Energy policy
it_topic17	International Affairs /Development policy
it_topic18	Law making process