Topic modelling and text analysis for qualitative policy research

Karoliina Isoaho, Daria Gritsenko, Eetu Mäkelä

University of Helsinki

Abstract: This paper contributes to a critical methodological discussion that has direct ramifications for policy studies: how computational methods can be concretely incorporated into existing processes of textual analysis and interpretation without compromising scientific integrity. We focus on the computational method of topic modelling and investigate how it interacts with two larger families of qualitative methods: *content and classification* methods characterised by interest in words as communication units and *discourse and representation* methods characterised by interest in the meaning of communicative acts. Based on analysis of recent academic publications that have used topic modelling for textual analysis, our findings show that different mixed-method research designs are appropriate when combining TM with the two groups of methods. Our main concluding argument is that topic modelling enables scholars to apply policy theories and concepts to much larger sets of data. That said, the use of computational methods requires genuine understanding of these techniques to obtain substantially meaningful results. We encourage policy scholars to reflect carefully on methodological issues, and offer a simple heuristic to help identify and address critical points when designing a study using topic modelling.

Keywords: Topic model, qualitative research, machine learning, big data, mixed-method research.

1. Introduction

Analysing text as data has been a significant component of policy science ever since Harold Lasswell's methodological contributions on content analysis techniques (Lasswell, Lerner, and de Sola Pool 1952). While later methodological developments have built on and drifted from the Lasswellian quantitative approach to focus on qualitative content analysis and discursive approaches to explore policy phenomena, policy scholars today are experiencing and having to react to a new, influential phenomenon, the so-called "computational turn" in social science. Access to data of unprecedented scale and scope has prompted a growing number of scholars to experiment with computational algorithms as the main or complementary method for text analysis – a laborious activity that has conventionally been executed manually by researchers (Mills 2017). These so-called *text-as-data methods* represent a large family of computational techniques (for an overview, see Grimmer and Stewart 2013). In recent years, the computational method of topic modelling (TM) has gained traction in policy-related studies. According to the Scopus database, TM is currently the most widely used text-mining technique in journals focused on policy analysis (Figure 1)[1]. Due to

---

[1]From 160 articles published between 2008 and 2018 using popular text-mining techniques (sentiment analysis, word embeddings, supervised learning, text clustering, text analytics or lda/non-negative matrix factorisation/topic model) 49% use topic modelling. This figure is based on the authors' key-word searches in the Scopus database (see Appendix 1 for full search queries).

this growing popularity, in this paper we put the use of topic modelling under scrutiny and examine its potential for qualitative policy research.
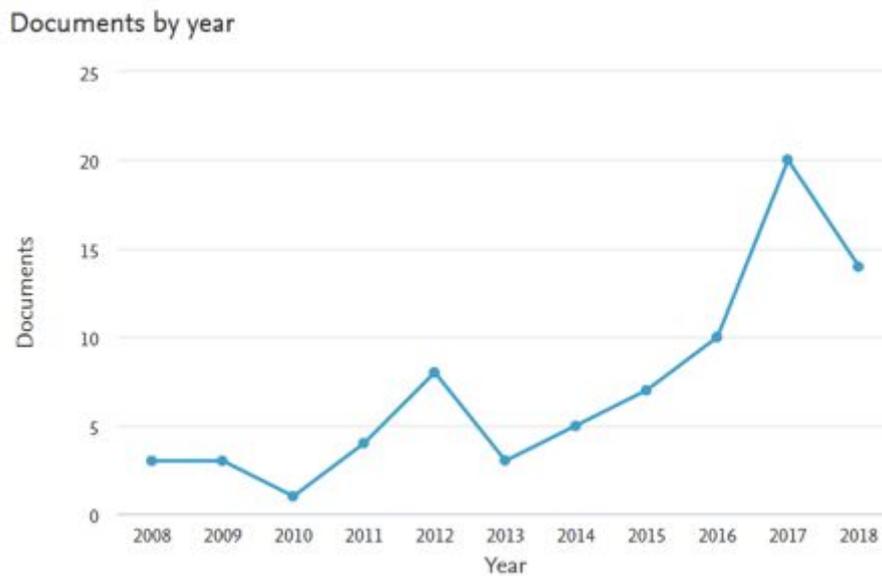

Figure 1. Number of articles using TM in policy-related journals in the past decade, 2008-2018. Based on the following query in the Scopus database: (ALL ("topic model*" OR lda) AND SRCTITLE (policy)) AND PUBYEAR > 2007 AND PUBYEAR < 2019

Topic modelling (TM) is a collective term for a family of computational algorithms that aim to "[discover] the main themes that pervade a large and otherwise unstructured collection of documents" (Blei 2012, 77). It has been argued that TM analysis is able to identify themes across large samples (Murakami et al. 2017) and to have "high levels of substantive interpretability" (DiMaggio, Nag, and Blei 2013). Consequently, the method's ability to "read" texts has been judged plausible in many cases (Mohr and Bogdanov 2013).

While these characteristics are appealing to scholars, the practices for applying TM have emerged sporadically. Automated analysis of text has been employed on a wide variety of different types of texts, including texts from social media platforms such as blogs, Facebook and Twitter (Dehghani et al. 2014; Kim et al. 2016); news media (DiMaggio, Nag, and Blei

2013; Grimmer 2010; Gritsenko 2016); and political and policy texts such as party manifestos, speeches, press releases, legislative proposals, and position papers (Isoaho, Moilanen, and Toikka n.d.; Munksgaard and Demant 2016; Quinn et al. 2006; Roth and Baumer 2016). While some scholars are paying more attention to the methodological dimensions of TM vis-à-vis textual analysis (Bryman 2006; Grimmer and Stewart 2013; DiMaggio 2015; Boussalis and Coan 2016), most applications feature no critical reflection on their research design – from corpus construction to result interpretation. To the best of our knowledge, there are no summarising accounts in English on the use of TM in qualitative research for policy research. Hence, there is no established convention under the conditions of which (if at all) TM can be useful for making sense of and interpreting policy text in social scientific research.

In this paper we seek to redress this gap by examining how policy scholars can take advantage of computational methods for textual analysis without compromising scientific integrity. The main aim of this paper is to discuss TM in the context of two groups of textual analysis methods: content and classification (C&C) methods characterised by interest in words as communication units, and discourse and representation (D&R) methods characterised by interest in the meaning of communicative acts (Titscher et al. 2000). The distinction between the two method families appears crucial to selecting the appropriate research design, as it conditions the novel methodological approaches and combinations at the level of both epistemology and strategy.

This paper is structured as follows. Section two introduces and discusses probabilistic TM and its variations. Section three compares TM with textual analysis methods to explore their

potential compatibility. In sections four and five, we discuss the use of TM together with text analysis methods in different mixed-method research designs and exemplify how these can be operationalised for policy research. Finally, section six sums up the discussion, also suggesting a simple heuristic to identify when using TM is appropriate in qualitative policy analysis.

2. Probabilistic Topic Modelling

2.1 Introducing topic modelling

Topic modelling is a collective term for a family of computational algorithms that are used to model text in a collection of documents as arising from a much smaller set of topics. These algorithms relate to latent variable discovery, dimensionality reduction as well as (soft) clustering. As they usually do not make use of human-curated seeds for the topics but instead derive them purely from the data, they are often classified under unsupervised machine learning methods. However, there are also extensions to topic models that do use seed words for topics (Jagarlamudi, Daumé III, and Udupa 2012), and the machinery of topic models has also been incorporated into classification algorithms that use teaching data (Mcauliffe and Blei 2008; Ramage et al. 2009), so this categorisation is by no means core to defining the method.

The first algorithm clearly identifiable as belonging to the family is Latent Semantic Indexing (Deerwester et al. 1990), which used the matrix factorisation technique of Singular Value Decomposition to derive latent variables from a set of documents. Because it was based purely on a linear algebraic transformation, LSI did not initially have a clear human-understandable explanation for how the latent topics were derived, hindering

interpretation and analysis. This was later rectified through framing the factorisation in probabilistic terms, which also led to a slightly improved probabilistic LSI model (Hofmann 1999), later shown to be equivalent to the most common form of nonnegative matrix factorisation, another linear algebraic technique (Ding, Li, and Peng 2008). Finally, in Latent Dirichlet Allocation (Blei, Ng, and Jordan 2003), the pLSI model was extended into an understandable, complete, *generative* probabilistic model. In practice this means that the algorithm contains a model for generating documents by randomly picking words from a set of topics based on probability parameters. Importantly, the probability parameters of the model can be tuned to better match evidence (i.e., an existing collection of documents) using Bayesian probabilistic inference. The tuned parameters can then be read back as a description of the topics in that collection (Blei 2012).

More specifically, the generative model encoded in modern TM algorithms is as follows: first, topics are modelled as bags of words, with a variable number of each individual word inside each bag (e.g., a particular topic may have many copies of words like "school", "teacher" and "degree", but very few copies of other words like "tree" or "kitten", so we could assume the topic is about education). The documents in a collection are also modelled as such bags, holding all of the words in the document, without regard to the order in which they appear. The TM algorithm then tries to recreate these document bags of words through the following process: First, from the set of topics covering the whole document collection, select some number of topics to which a particular document pertains, and their proportions in the document (e.g. this document is 67% about education and 33% about environment protection). Then, recreate the document word bag by sampling words at random from each topic bag in the thematic proportions selected previously (Blei, Ng, and Jordan 2003).

When the TM algorithm begins to run, the proportions of words contained in each topic bag are initialised as random, as are the proportion of topics in each document. Then, with multiple rounds of Bayesian inference, both proportions are gradually changed, so that the bags of words that the model produces correspond as much as possible to those derived from the actual data.

2.2 Topic output

The output of TM contains two items: the topic word proportions and the document topic proportions. At the end of the training process, both output items can be read out of the model and subjected to human analysis. Figure 2 shows a typical display in which topics are presented as a set of word lists containing the top five words associated with each topic, and documents are displayed by showing the top four topics associated with them, along with their proportions in the document.

Topic	5 top terms
0	conservation variety graywater employee onboard
1	tui ag development report group
2	ships safety cruise ship training
3	shipboard employees company programs waste
4	management environmental water system environment
5	princess passengers cruises alaska onboard
6	carnival lines cruise environmental guests
7	waste guests order means economic

8 sustainable energy employees social german

9 carnival uk en la cunard

#doc	name	topic	proportion	topic	proportion	topic	proportion	topic	proportion	topic	proportion
0	file:/C:/data/file1.txt	7	0,47	4	0,28	9	0,23	3	0,01		
1	file:/C:/data/file2.txt	8	0,38	9	0,36	4	0,22	3	0,02		
2	file:/C:/data/file3.txt	9	0,47	3	0,32	4	0,18	0	0,04		
3	file:/C:/data/file4.txt	6	0,42	4	0,29	9	0,27	3	0,02		
4	file:/C:/data/file5.txt	2	0,50	9	0,35	4	0,12	8	0,01		
5	file:/C:/data/file6.txt	9	0,52	0	0,22	4	0,15	3	0,11		
6	file:/C:/data/file7.txt	9	0,41	5	0,34	4	0,17	3	0,05		
7	file:/C:/data/file8.txt	1	0,52	4	0,36	9	0,11	8	0,01		

Source: Authors.

Figure 2. Topic modelling output, from the Mallet software. Analysis based on annual sustainability reports issued during 2008-2013 by Carnival Corporation and Royal Caribbean Cruises.

How well these derived topics correspond to any phenomena of interest to a researcher depends on 1) how well the collection of documents can be thought of as having been created by the model described above (each document deals with some number of topics in some proportion; the topics talked about determine the vocabulary used), 2) how well the definition of a topic in the above model corresponds to the phenomena of interest, and 3) various

assumptions about the exact topic model variant used that further define how the topics behave.

As examples of the effect of such assumptions, the original Latent Dirichlet Allocation (LDA) -based topic model sampled both the proportions of words in topics as well as topic proportions for documents from symmetric Dirichlet probability distributions. A symmetric Dirichlet distribution over document topic proportions is well able to represent: 1) a collection of documents in which each document mentions all the topics, 2) a collection of documents in which most documents mention a few topics, and 3) a collection of documents in which the number of topics mentioned varies widely between documents. However, it is not able to model a collection well when a particular topic is consistently and widely present, while others appear together only rarely (e.g., a collection on various branches of EU policy in which EU terminology is consistently present, but other topics vary widely). In effect, the symmetric Dirichlet distribution is not well suited to modelling a selection of topics that do not all behave the same way.

This does not mean that the topic model does not work at all for such materials, but that the resulting topics will be worse. Because of this, traditional LDA has required a significant amount of pre-processing to remove general language words (such as "the", "and", etc., commonly referred to as 'stopwords'), as these confuse the model. On the other hand, when the requirement for topics to behave identically is loosened in asymmetric LDA (Wallach et al., 2009), the model is able to segregate these general language words into a single topic without adversely affecting the quality of the other topics.

Similarly, traditional Dirichlet-based LDA models assume that topics appear in documents independently of each other. As this assumption clearly does not hold in most collections of texts in practice, extensions of the LDA have been developed. The Correlated Topic Model (CTM) (Blei and Lafferty 2006) replaces the topic proportion prior with one capable of capturing correlations between topics (so that, for example, the topic "mantle isotopic crust plate earth" often appears with the topic "fault earthquake data earthquakes images"). However, while this improves the ability of the model to mimic the original texts, as well as providing topic correlation numbers as possibly useful information for the researcher, it has been reported (Chang et al. 2009) that the topics generated are themselves less interpretable by humans. This may arise from the fact that correlating the topics means that they are no longer as independent and distinct as possible.

Another direction for extending topic models has been the inclusion of text-external correlates into the model. Dynamic Topic Models (Blei and Lafferty 2006) add time as such a correlate, enabling charting and comparing, for example, how a certain topic is discussed in a collection at different times. Other extensions (Mimno and McCallum 2008; Rosen-Zvi et al. 2010) add categorical correlates, allowing comparison by, for example, author, group or political leaning. As a final useful extension to TM, the Structural Topic Model (STM) (Roberts et al. 2014) combines the correlated topics of CTM with the text-external correlates from DTM and other attribute-correlate models. STM also provides tools for statistical model evaluation, including heuristics for selecting an optimal number of topics to be extracted.

Finally, it must be noted that, in practice, before a topic modelling algorithm is run, its input often undergoes substantial pre-processing, aimed at unifying the vocabulary and making it

more amenable to topic deduction. Besides the already mentioned stopword removal, steps taken often include removing all punctuation, replacing all inflected words with their base forms, replacing all numbers with a single tag, and unifying the same words written with different capitalisation. While in a strict sense outside the application of the topic modelling algorithm itself, these pre-processing steps may, however, have a great effect on its results, and thus should also be subjected to rigorous examination and validation (Denny and Spirling 2018). This is particularly important due to recent research which shows that many commonly accepted pre-processing steps have no or even adverse effect on topic modelling (Schofield, Magnusson and Thompson 2017a, Schofield, Magnusson and Thompson 2017b).

2.3 Model validation and interpretation

When topic model is run, three intertwined steps are required to ensure the validity of the analysis. First, the output of the model needs to be interpreted. Second, the choice of pre-processing and modelling parameters needs to be validated. Finally, the ability of the topics to model the phenomenon under investigation needs to be evaluated.

Unfortunately, many studies currently forgo rigorous evaluation as well as employing flawed interpretation practices. First, it is a common practice to interpret topic output in isolation from the documents, and based on only the top five to twenty words most associated with the topic in the collection. One might argue that this practice is flawed, because the topics are a description of the collection of documents, and do not exist in isolation from it. Thus, they also need to be interpreted not in isolation, but within the context of the documents in which those words originally appear. In addition, summarising with just the top N words what is in

reality a distribution over all the words appearing in the textual material can hide important and interesting information.

Besides misrepresenting actual content, this practice may hide problems in model parameterisation. For example, if the number of topics or the preprocessing parameters have been incorrectly set, a generated topic may end up as an amalgam of multiple distinct topics. Yet, particularly given the capability of the human mind to find connections everywhere, this may not be apparent from just the top five words in the amalgamated topic. In short, interpreting short word lists in isolation paves the way for misinterpretation of what the topics identified signify in the documents.

At the same time, there is ample prior research supporting rigorous validation of topic model output. DiMaggio et al. (2013, 586) identify three forms of validation; statistical, semantic, and predictive. Of these, statistical validation usually starts with an analysis of how well the model can predict a held-out set of documents. By comparing statistical perplexity estimates between models run with different parameters, better values for the number of topics as well as other parameters can be found. Beyond searching for the best parameters, statistical validation can also be used to evaluate how well the data fit the statistical assumptions in the model in general (Mimno and Blei 2011). Tang et al. (2017) have also used statistical measures and theoretical analysis to explore the limits of topic modelling with regard to the number of documents and their lengths. From their experiments, it appears that topic modelling requires documents of at least 100–200 words in length, while the number of documents needs to be at least 1000–2000. In addition, topic modelling works better if the sets of vocabulary associated with the different topics are clearly separated. The quality of the

output also depends crucially on the researcher selecting the right number of topics, further highlighting the need to validate the value of this parameter rigorously.

At the same time, as evidenced by the lower interpretability of correlated topic models despite their lower perplexity (Chang et al. 2009), statistical measures should not be taken blindly as the sole method of validating a model and its parameter choices. Instead, proper validation requires a combination of statistical, semantic and external measures (DiMaggio 2015). In semantic validation, the researchers reflect on the model output to evaluate its interpretability and plausibility. A rigorous way to do this, which Boussalis and Coan (2016, 94) call concurrent validation as distinct from semantic validation, is to manually code a sample of the data and compare these results with the model output.

Finally, in external validation, researchers verify that the model reflects relevant collection-external information. For example, topic prevalence can be verified to respond to relevant collection-external events such as the news attention cycle or political debate schedules.

As examples of how these different forms of validation and interpretation can feed beneficially into each other, Boussalis and Coan (2016) improved the semantic validity of their results by using statistical metrics to examine the "semantic" distances between topics. Similarly, Mimno and Blei (2011) used measures of statistical deviation from model assumptions to aid in interpretation of the topics, as well as the effects of various coefficients.

The usefulness of these approaches is always contingent on the research aims of each individual study, and therefore researchers have to evaluate how each validation measure can be used in their own work (Boussalis and Coan 2016). Nonetheless, the key rule of thumb

applies to all scholars wishing to use topic modelling in text analysis: "validate, validate, validate" (Grimmer and Stewart 2013). In the words of Grimmer and Stewart (2013, 5), which are often cited but deserve to be repeated:

"(T)he output of the models may be misleading or simply wrong. Therefore, it is incumbent upon the researcher to validate their use of automated text analysis. (…) What should be avoided, then, is the blind use of any method without a validation step."

3. Application of TM in social science research

In the extant literature, TM has often been equated with multiple textual analysis methods commonly applied in social science research. Yet, as indicated in the Introduction, it is useful to distinguish between the methods interested in words primarily as communication units and those focusing on the meaning of communicative acts. This distinction allows us to discuss TM more carefully vis-à-vis the groups of C&C and D&R methods, taking into consideration their epistemological and practical considerations. Table 1 summarises and compares the characteristics of each method considered in our analysis[2].

Group of methods	Type of analysis	Unit of analysis	Study starts with...	Subjectivity
Content and classification	Inductive Content Analysis/Grounded theory	code-category-concept	observation	aims at objective and systematic result
	Deductive Content Analysis	code-category-concept	theory, previous research	

---

[2] This is not intended to be a comprehensive or exhaustive list of all textual analysis methods. Detailed information on other approaches to text analysis that could be used for content and classification or discourse and representation can be found, for example, in Denzin (2007), Kuckartz (2014) or Silverman (2016).

	Thematic	code-code family-theme	familiarisation with data	
	Vocabulary	word-to-word and word-to-concept co-occurrences	(key)words	
Discourse and representation	Frame analysis	framing devices (catchphrases, metaphors, sound bites, graphics, and allusions to history, culture, or literature)	thorough reading of content	research-generated knowledge; subjectivity of researcher cannot be avoided
	Critical Discourse Analysis	hidden relations of power present in text; agency, contextual factors, language used to characterise events and people	observation, scoping phase to get an understanding of the material	
	Argumentative Discourse Analysis	storylines, discourse coalitions	identification of key passages of argumentation, identification of storylines	
	Narrative analysis	story	reading text, compiling stories	

Table 1. Comparative table of text analysis methods. In this paper we follow Titscher et al. (2000) to categorize them as either *content and classification* methods (characterised by the interest in words in single, compound or idiomatic terms) or *discourse and representation* methods (characterised by the interest in meaning in communicative acts such as discourses).

As Table 1 shows, the "analysis pipeline" in C&C methods is closer to the process of TM, while D&R methods diverge more significantly from the use of TM. As a result, while TM is increasingly applied in social science research, current attempts to understand TM and its output in textual analysis vary, and risk being misleading. Even if Blei originally proposed TM as a tool to provide a "browsing experience" or as "an algorithmic solution to managing, organizing, and annotating large archives of texts" (Blei 2012, 77-79), TM is increasingly

used as a method to reveal aspects of discourse environments (DiMaggio, Nag, and Blei 2013; Goldstone and Underwood 2014; Munksgaard and Demant 2016), semantic or thematic categories (Jaworska and Nanda 2016; Mohr and Bogdanov 2013), issue definitions (Nowlin 2016), narratives (Grubert and Algee-Hewitt 2017), frames (DiMaggio, Nag, and Blei 2013) and author characteristics (Seroussi, Zukerman, and Bohnert 2014). Clearly, these interpretations lack common coherence and raise the question of whether output from TM can represent such varied phenomena and concepts. This is a critical methodological question because, as we explained in sections 2.1 and 2.3, one of the determining requirements for making topic models interpretable, in addition to guaranteeing model robustness and validity, is making the topics correspond to the phenomena of interest.

The output of theory-driven D&R methods in particular includes contextual and semantic understanding. It has been argued that the data-driven TM has limited ability to grasp these aspects because it only analyses words contained in the document corpus. As Klein et al. (2015, 132) aptly put it, "(t)here is no inherent reason to believe that words grouped together on the basis of co-occurrence statistics should really mean or prove anything, aside from the winkingly suggestive similarities that these word groups so often display. For topic models to be truly integrated (…), the user must be able to probe the semantic associations that model proposes, and seek out additional perspectives on the model itself". Indeed, while algorithmic exploration of context is currently being developed and studied, it is not given that TM, in its current state, lends itself to examining the hidden power relations, agency and context that are at the core of D&R methods.

In addition, subjectivity is explicitly acknowledged in D&R methods. The discursive storyline, frame or narrative story that emerges from a text analysis is always formulated by a researcher who has a specific question in mind and an understanding of the implicit context and background-knowledge factors that the readers of a given text are assumed to be aware of (Hajer 1995; Schön and Rein 1994). TM and C&C methods are both based on positivist assumptions and the analysis process is grounded in text with manifested items (words, sentences, paragraphs, documents) in focus. However, TM provides a generative model, while content or thematic analysis aim at "the subjective interpretation of the content of text data through the systematic classification process of coding and identifying themes or patterns" (Hsieh and Shannon 2005, 1278). It follows that, rather than claiming straightforward qualitative value from TM analysis, the potential of TM may be best harnessed in mixed-methods designs.

4. Topic modelling in mixed-method designs

We now discuss two mixed-methods approaches, embedded and sequential design, and give examples of good practices in applying TM to textual analysis. To substantiate our methodological argument, we report on 25 recently published studies that have applied TM to conduct textual analysis in a qualitative setting. These studies were identified using the Scopus database[3] and reviewed by the authors (Appendix 2). A summary of the considerations and limitations of using TM with C&C and D&R methods is presented in Table 2 at the end of this section.

---

[3] TITLE-ABS-KEY ("topic model*" OR "LDA" AND text*) AND (LIMIT-TO (DOCTYPE , "ar") OR LIMIT-TO (DOCTYPE , "ip")) AND (LIMIT-TO (SUBJAREA , "SOCI")) and hand-picking for relevance to policy research.

4.1 Embedded design

An embedded design means that a project has a predominant method that guides the analysis and an auxiliary method used to enhance this process (Creswell 2003). The benefit of this design is that a researcher may gain a broader perspective on the problem, use different methods to study different levels within a design, or treat different aspects of the materials differently (see Creswell et al. 2003, 230). We argue that content, thematic and vocabulary analysis can be combined with TM through applying embedded design (Figure 3).
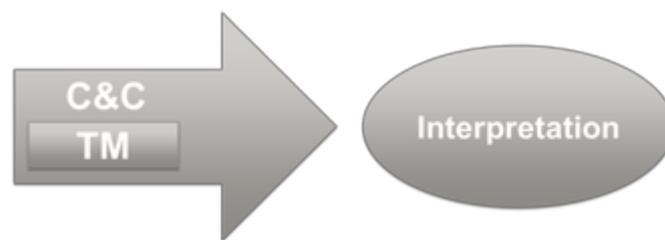

Figure 3. Content and classification methods and topic modelling in an embedded design.

4.1.2 Content analysis

Content analysis can be usefully complemented with TM techniques. Baumet et al. (2017) argued that TM is surprisingly similar to grounded theory, a popular inductive implementation of content analysis, demonstrating how similar results could be obtained by applying these two methods to the same research problem. They saw the main similarity at the "strategy" level – that is, a provisional theory (in case of grounded theory) or model (in case of TM) that is iteratively refined based on the data. However, we argue instead that GT is dissimilar to TM precisely at the level of strategy: GT aims at a theory (an explanation of certain outcomes) while TM provides a model (an examination of certain variables). They do have similarities, though. First, both methods are iterative. The constant comparison and revision of coding categories in qualitative content analysis is similar to the topics iteratively

coalescing through Bayesian belief update operations in TM. Second, in both methods the same word could be included in more than one category. Yet, unlike grounded theory, TM takes all the words in the text into account, which creates a reliable and replicable way to deconstruct the whole collection of texts.

The following strategies for embedding TM into content analysis can be used. Researchers may consider TM output during the process of initial coding, a procedure to sift through large volumes of data in a systematic fashion, often implemented in the framework of grounded theory methodology (Glaser and Strauss 1967). It has been demonstrated that grounded theory with a "computational twist" can utilise TM at the exploratory stage, deep reading at the pattern refinement stage, and supervised machine learning approaches at the pattern confirmation stage to provide a robust, rigorous, and reproducible methodological framework (Nelson 2017). While grounded theory explicitly aims to develop small or mid-range theories and holds prescriptions for sampling and quality control, the same steps can be used in other contexts that benefit from the inductive approach to content analysis (Boussalis and Coan 2016; D 'Amato et al. 2017).

4.1.3 Thematic analysis

Thematic analysis is usually described as "a method for identifying, analysing and reporting patterns (themes) within data" (Braun and Clarke 2006). It is compatible with various theories, it can be inductive or deductive, and it employs coding as a technical procedure for working with the data. Thematic analysis does not attempt to quantify themes, calculate their frequencies, or establish any other proxy for significance (Vaismoradi, Turunen, and Bondas 2013), and it is often used in the initial stages of a research project with a large dataset, while

interpretative methods can later be applied to its specific parts and in response to more specific research questions.

TM has been used as a form of thematic analysis to estimate the salience of particular issues within the multidimensional policy space (Nowlin 2016). Klein et al. (2015) combined TM (which they conceived of as automated thematic analysis), sense-making, and interactive visualisation in their TOME (interactive topic model and meta-data visualisation) model. Their work illustrated the usefulness of TM for this task. Similarly, Greene and Cross (2017) used TM to reveal the themes in EU parliamentary debates. Statistical regularities of natural language seem to present quite an accurate picture of content structure. Nevertheless, it is important to emphasise that the topics generated by the model might also identify specific historical events, notable stylistic features, or systematic transcription errors, to name only a handful of non-thematic topics.

The central challenges pertaining to thematic analysis are defining what a theme is and how many themes there are in a given dataset. TM not only allows exploration of a much larger corpus, but also the discovery of themes in a more systematic way. Thus, there is potentially a high degree of convergence or even potential for substitution of thematic analysis with TM if the researcher wants to know what a dataset is about (Murakami et al. 2017). Both methods are grounded in text and their output is an overall description of the dataset obtained by decomposing it into separate themes. Neither of them pays attention to language use or agency (i.e., who is speaking). However, such an integration requires that interpretation of a topic takes into consideration not only the top N words, but also the top documents in a topic. Only by inspecting the documents associated with a topic can a researcher capture a theme,

defined as the underlying meaning within the dataset that captures something important about the data in relation to the research question (Braun and Clarke 2006). While TM can be embedded into TA, the latter cannot be a purely data-driven approach, since sampling, curation of the corpus, and interpretation of topics in relation to the corpus are paramount.

4.1.4 Vocabulary analysis

Vocabulary analysis is a group of techniques that focuses on "the system of words and their meanings commonly used by social collectives" (Loewenstein, Ocasio, and Jones 2012, 41). Its main assumption is that "meanings do not reside in words but rather in how words relate to their linguistic environment – that is, how words relate to other words" (Krippendorff 2004, 290). Most commonly, vocabulary analysis is associated with identification of word-to-word and word-to-example relations, i.e., a network of relationships within a text.

Word-to-word analysis – sometimes termed map analysis or the analysis of situated concepts – seeks to elaborate networks consisting of connected concepts rather than frequency counts or underlying themes (Carley 1994). In practical terms, a researcher seeks to determine whether certain words occur in relation to other words in a set of texts. "Differences in the distribution of concepts and the relationships among them across texts provide insight into the similarities and differences in the content and structure of the texts" (Carley 1993, 92). Word-to-word relations in the resulting "maps" can be examined both graphically and statistically (Carley 1993). Therefore, word-to-word vocabulary analysis based on word co-occurrence is very close to TM in its idea, although their technical implementations differ. Carley (1993, 102) noted that "a disadvantage to map-analytic techniques is that they are harder to automate". In those early days of computer-assisted text mining, automation was

already applied to the coding of concepts, but the coding of relationships that is required for word-to-word vocabulary analysis had not yet been automated. TM is a tool that has automated the coding of relationships between words by assigning them probabilities of belonging to the same topic, as well as providing other metrics of interest, such as distances between topic distributions.

Word-to-example analysis seeks to address the limitations of word-to-word analysis with regard to problems concerning symbol grounding and meaning holism (Loewenstein, Ocasio, and Jones 2012). One type of such analysis – the *practice perspective* developed by Mohr and Duquenne (1997) – examines the co-occurrence of words and practices. Other scholars have studied proper names (firm names (Kennedy 2008), program names (Nigam and Ocasio 2010)) in relation to vocabularies. This approach showed what best exemplified certain categories in a dataset and, vice versa, which vocabularies were used to describe certain examples. We suggest that STM could be used for word-to-example analysis as it allows the use of background variables to discriminate between topics (Chandelier et al. 2018; Lucas et al. 2015).

Thus, at least in theory, TM can be embedded in vocabulary analysis since it allows the representation of a collection of texts as a collection of vocabularies, and inspecting the parts of the documents that include top words provides a key-words-in-context (KWIC) view of a derived topic. Dynamic topic models even offer the opportunity to study how vocabularies evolve over time – a task that has been notoriously laborious to perform through hand-coding. As we have not encountered this kind of mixed-method setting at the time of writing this article, the design requires further testing. The difference between the vocabulary

analysis and TM methods resides at the level of research strategy: while an analyst performing vocabulary analysis may want to focus on specific vocabularies, TM algorithms provide a large number of vocabularies in the form of topic word proportions that may or may not be related to the topic of research. However, we argue that the assumption of vocabulary analysis – that systems of words that belong together have special properties (namely, words are context for other words) allowing researchers to bridge the gap between words as explicit lexical units and their implicit meanings – can be extended to TM. This makes the avenue of enhancing vocabulary analysis with TM look most promising.

4.2 Sequential design

A sequential design refers to analysing first qualitative and then quantitative data – or vice versa – to understand a given research problem. In this design, data is analysed separately and in consecutive phases of the research process (Creswell 2003). While the qualitative analysis takes precedence, the quantitative results are important in guiding the process of qualitative analysis. The benefit of a sequential design is that it allows the two methods to complement each other and contributes to a more robust data analysis process.

Figure 4 demonstrates how TM can be combined with D&R methods in a sequential design. Given that the application of D&R methods rarely follows a linear path in practice but includes numerous rounds of iteration between the data and the emerging interpretation, there is room to apply TM to guide and inform this process. We now consider ways of using TM and different D&R methods in a sequential design.

Figure 4. Discourse and representation methods and topic modelling in a sequential design.

4.2.1 Discourse analysis

Discourse analytical methods explicate how reality is socially constructed through knowledge production and meaning making (Keller 2013). While, as we explained in section 3, researchers should be wary of interpreting topic output straightforwardly as discourse due to the strong theoretical basis of discourse methods, topic models can help in revealing aspects of discourse when used in a sequential setting.

For example, several researchers have used TM to guide the first steps of discourse analysis, which often includes scoping the data and conducting first rounds of coding to get familiar with the data. Törnberg and Törnberg (2016a, 2016b) used TM in combination with Critical Discourse Analysis to examine discursive connections between Islamophobia and anti-feminism in online discussions. The authors used TM in the first phase of their research to obtain inductive empirical categorisations of their corpus for the subsequent discourse analysis. They found this sequential design enriching, as TM inductively revealed passages where certain discursive fields were present (Törnberg and Törnberg 2016a, 133). Similarly, Light and Cunningham (2016) and Lindgren et al. (2018) used a sequential two-stage approach to examine media discourse. In both studies, TM was first applied to provide an overview of the corpus. Light and Cunningham used TM in examining how international peace movements were discussed in the Nobel Peace Prize discourse across decades. TM was used to locate themes in the Nobel speeches and the topic output served as "content codes",

which were then opened up in a qualitative discourse analysis of the text. Hence, TM can inform (and to some extent automate) the first steps of discourse analysis, such as examining common collocates to certain words and then putting them into thematic categories.

In addition to informing and guiding the analysis process, applying TM in discourse analysis can add analytical rigour to the analysis. Törnberg and Törnberg (2016a) and Jaworska and Nanda (2016) found that applying TM leveraged the selection bias related to coding of text – one of the critical issues related to applying these methods. Another key benefit of TM relates to its ability to list the documents most strongly associated with certain topics. This makes it possible for researchers to identify a smaller set of documents for the subsequent analysis of discourse, which allows them to go past the topic structure and account for the intertextual and contextual factors that go beyond the analysis of words (Lindgren 2018). In sum, as a standalone method, TM is limited for the purposes of conducting a discourse analysis. However, the method can provide useful insights when it is applied separately and when its limitations are taken fully into account.

4.2.2 Frame Analysis

Frames are commonly defined as "coherent interpretative packages" (Entman 1993; Gamson and Modigliani 1989). In this view, the patterns, elements and assumptions that emerge around a certain theme constitute a frame. Once identified, frames can be further analysed by interpreting their meaning or weighing their effectiveness. The ability to automate frame identification has made many scholars explore whether TM could ease the analytical process in frame analysis (Jacobi, van Atteveldt, and Welbers 2016; Pashakhin 2016). The authors turned to TM, arguing that the conventional content and discourse analysis methods often

used for frame discovery are limited when facing a large amount of text and are also prone to subjectivity both in text selection and analysis.

However, while the benefit of scale and scope of a TM analysis is clear, combining TM with frame analysis poses a number of issues. Van Atteveld et al. (2014) and Jacobi et al. (2016) examined whether topics can be seen as applied frames and concluded that the topic output does not approach the methodological definition of frames as coherent interpretative packages. Instead, some topics are connected more with substance than others and topic interpretation depends on the case and research question in question. Moreover, TM does not grasp a crucial part of frame analysis, that is, considering what is *not* represented in the frames and what is left missing in the text when adhering to a fixed frame.

Nonetheless, TM can provide useful empirical avenues for frame analysis. Some topics "are very high precision indicators of issues", which "opens up possible strategies to use selected topics as substantive indicators" (Atteveldt et al. 2014). Jacobi et al. (2016) combined TM, sentiment analysis and qualitative frame analysis to provide better results of frames and intertextuality. In sum, the groundwork for frame discovery can be initiated by applying TM and the empirical usefulness of TM is that it allows for the analysis of temporality and continuity of topics. Yet, it is imperative to note that TM remains an incomplete standalone method for frame analysis.

4.2.3 Narrative analysis

Narrative analysis refers to techniques united by their attention to the story as an object of investigation. Narrative can refer to a life story, extended account of lives and events, and

stories organised around characters, setting, or plot (Riessman 2005). The method seeks to emphasise sequence and consequence in texts, rather than picking up snippets of text detached from their broader context (Mishler 1995). As the role of researchers in narrative analysis is substantial in distilling the narrative from text, building upon the implicit context and background-knowledge, TM seems methodologically very distant from narrative analysis.

However, there are studies that have suggested using TM in narrative analysis. Jaworska and Nanda (2016) as well as Grubert and Algee-Hewitt (2017) propose to interpret topic output as signals of narratives or the latent foreground for narratives. First, they argue that the ability of STM to leverage meta-data from documents could make it useful for revealing narrative themes and linking these to information on document authors. In this way, STM could provide a starting point for unveiling latent narratives, and also for identifying characters (agents) who belong to the narrative story.

In addition, dynamic topic models such as STM make it possible to trace the temporal evolution of topics in a corpus. This kind of analysis could be used to gain knowledge of shifts in meta-narratives and the narrative sequence structure. The STM analysis could assist researchers in identifying the beginning, middle and end of narratives. Subsequently, the STM output can be used to guide more detailed analysis on narrative structure. However, the application of STM is useful only if the time-frame covered by the data is clearly identified. Importantly, TM output should not be interpreted as coherent narratives. We see specific potential for STM to act as complements to narrative methods.

	Method	Integration with/of TM	Considerations and limitations
High degree of convergence or substitution with TM	Content analysis	Use embedded design. Iterative method to compare and revise coding categories.	Some content analysis methods, like grounded theory, are dissimilar to TM at the level of strategy. GT aims at theory, TM is a model.
	Thematic analysis	Use embedded design. Mining of themes in a systematic way.	Topics generated by the model can also refer to non-thematic issues. Investigating documents associated with the topic is important to validate results.
	Vocabulary analysis	Use embedded design. Vocabulary analysis based on word co-occurrence is very close to TM in concept, although their technical implementation differ.	Differ at the level of strategy: Vocabulary analyst often focuses on specific vocabularies, while TM method includes a wide range of words that may or may not be relevant to the topic of research.
No or little potential in substitution of method with TM	Discourse analysis	Use sequential design. Generate directions for analysis of discourse, hidden power relations, agency etc. TM can provide a systematic sampling method for discourse analysis.	Discourse methods have a strong theoretical basis, and take into account contextual, intertextual and semantic factors, which TM cannot.
	Frame analysis	Use sequential design. Provide empirical avenues for frame analysis and aid in frame discovery.	TM does not take into account contextual, intertextual or semantic factors that are integral in frame analysis.

| | Narrative analysis | Use sequential design.

Topic output could provide a starting point in revealing meta-narratives, sequence structure in narratives or characters belonging to a narrative. | TM does not take into account context or background knowledge. |
|---|---|---|---|

Table 2. Potential and limitations of TM for C&C and D&R methods.

5. Applications for policy studies

Finally, we draw attention to three studies that have used TM with policy concepts and theories. In doing so, we do not attempt to cover the links exhaustively, but give an insight into how TM can be applied to research policy concepts.

Nowlin (2016) is an example of a study that benefitted from an *embedded design*, substituting thematic analysis with TM. The paper applies LDA to a large collection of Congressional hearings to generate a model of issue definitions. This model suggests that issue definitions are multi-dimensional, that is, have multiple attributes, but "only certain attributes (i.e., dimensions) of an issue may be salient at any point in time" (Nowlin 2016, 312). When actors manipulate the dimensions of issue definitions they frame (or reframe) an issue in a particular way. Over time, such framings aggregate and become collective issue definitions. Hence, to study "issue definitions" one needs to identify the possible "dimensions" of an issue, when and how they were deployed, and how these processes evolved. This could be done by performing thematic analysis – identifying themes vis-à-vis a given research question (in this case, a qualitative researcher would have asked "Which characteristics are associated with the policy issue in question?" and coded only those parts of the materials relevant to answering this question). Nowlin's (2016, 320) study assumed topics to be "synonymous with the

dimensions of an issue". Issue salience was operationalised through topic prevalence in the documents. The practical steps of this research were:

1) constructing a corpus dealing with one specific issue (used nuclear fuel) over the period 1975-2012;

2) applying LDA to this corpus;

3) interpreting topic output as "terms most associated with each dimension";

4) labelling each dimension and observing the topic proportion over time.

The author reflected on the corpus construction in depth, demonstrating how the chosen material (opening statements of witnesses appearing at Congressional hearings) is representative of the characteristics these actors attached to the problem of used nuclear fuel. Result validation was performed by placing the findings within the context of "exogenous events", namely, changes in the legislation. A simple OLS regression with the legislative acts as independent variables was used to examine the mean proportion of each topic in the corpus, aggregated by year. The study concluded that the approach successfully captured the shifting issue definitions. While this is an interesting approach, the author reflected on its central drawback – the inability of LDA to model a selection of topics that do not all behave the same way based on the assumption of symmetric prior (see Section 2.2). The study refers to this drawback as missing "some of the nuance associated with particular attempts to reframe a policy issue" (Nowlin, 2016, 326), but as we argued in this paper, it represents the methodological problem of a mismatch between the phenomenon of interest and the technical assumptions of the LDA algorithm.

Therefore, in line with the multi-faceted validity criteria for TM, an embedded topic modelling design for policy analysis, such as the above study, will benefit from relying on the original documents for the interpretation of the topic output. In addition, since LDA assumes that the same topics are present throughout the whole corpus, researchers should in general be wary of using the method to capture the emergence of new topics – which could instead be leveraged by using algorithms with relaxed assumptions, such as STM. Interestingly, the work by Gilardi, Shipan, and Wueest (2017) provides an example of investigating issue definitions using STM. In examining restrictions on smoking across U.S. states, the authors used topic prevalence within a state as their measure of issue definition. They first identified topics and their distribution over time, drawing on a dataset of more than 3 million paragraphs that focus on anti-smoking laws in 49 states, which was followed by four validity checks for the topics (these include sentiment analysis and considering how topics correlate with the timing of the adoption of smoking bans at state level). Finally, the authors engaged in conceptual discussion, arguing that "topics varied with the frequency of earlier adoptions by states within the diffusion network" and that "topic prevalence is linked to prior policy adoptions within a state's diffusion network" (Gilardi, Shipan, and Wueest 2017, 13), shedding light on whether and how the issue definition stage of the policymaking process is linked to prior adoptions. Although the setting is not directly comparable with Nowlin's work, it shows one clear benefit of using STM over LDA. Giraldi et al. incorporated several covariates into their model: this enabled them to examine which topics appear most frequently, which have remained stable over time, and other variations in topic prevalence. However, for such an approach to work, the policy must be widespread, allowing for the creation of a corpus of sufficient volume.

Another example is the *sequential* application of TM in frame analysis by Fawcett et al. (2018). The aim of the study was to explain why issues "fade" from the problem stream, thereby improving our knowledge on how "problem windows" form and change. The authors applied TM together with conventional hand-coding to track the co-evolution of frames and frame sponsors, i.e. powerful elite actors who impact problematisation. To do this, they used a sequential design with the following four steps:

1) The relevant data was captured and prepared for the topic modelling.

2) The TM analysis was run, the authors selected a 15-topic model and one policy-related topic from within that model.

3) Four problem windows were identified from within the selected topic.

4) Finally, the presence of frames and frame sponsors were hand coded within the news articles published during the four problem windows identified in Step 1 (Fawcett et al. 2018, 7).

With such a sequential design, Fawcett et al. were able to identify relevant time periods and zoom in on those to then conduct further qualitative analysis on news articles published at a specific time. They argue convincingly that such research design has benefits over the conventional method of using "article frequency count as a justification for 'zooming in' on particular periods of time" (Fawcett et al. 2018, 6). TM allowed them to select articles from the actual periods of problematisation, rather than those stemming from "the other types of extraneous reporting" (Fawcett et al. 2018, 6).

6. Conclusion

This paper set out to discuss the contribution of TM to textual analysis in a qualitative setting. The consequent aim was to open up the logic behind TM to policy scholars and highlight that the scholarly capacity to craft robust research designs for TM depends on researchers' familiarity with the technique, technical assumptions, and good knowledge of the phenomena being investigated.

The first major point about TM presented in this paper is that that its output contains *two* items: the topic word proportions (usually presented as a set of word lists) and the document topics proportions. We highlighted that the extent to which these derived topics correspond to any phenomenon of interest to a researcher depends on: 1) how well the collection of documents can be thought of as having been created by the generative model underlying TM, 2) how well the definition of a topic in that model corresponds to the phenomena of interest, and 3) various assumptions regarding the exact topic model variant used that further define how the topics behave. All of the above need to be known, understood and validated before a TM analysis can be trusted.

With the above definitions outlined, we discussed TM together with the analytical procedures of commonly used C&C as well as D&R methods. Regarding C&C methods, as the TM algorithm can automate the process of analysis, it can, depending on the method, either completely (as with thematic analysis) or partially (as with inductive content analysis and vocabulary approach) substitute for procedures previously performed 'by hand'. This is possible thanks to TM sharing similar positivist assumptions, being grounded in data and focussing on text as a manifestation of explicit meaning. Such substitutions should, however, not be done without critical considerations of methodological objectives. We encourage scholars to further experiment empirically with TM to test the extent to which TM can

automate C&C processes. The extensions of the TM method that overcome some of the limitations of the original LDA algorithm seem particularly suited to this purpose.

With regard to TM and D&R methods, the first determining finding is that the output of TM should not be equated with discourses, frames, or narratives. The latter concepts are highly informed by theory, communicating contextual, representational and semantic meaning, which are not currently within reach of the Bayesian-probabilistic-based algorithm. As such, our heuristics suggest that TM cannot substitute for any part of D&R methods in terms of analytical procedures due to the divergent epistemological considerations. Nonetheless, while future developments of TM may allow the algorithm to be trained to inform about discourses, frames and narratives, we argue that in its current state TM can add value when used as a *complement* to D&R methods. TM can be integrated in the analysis sequentially as part of mixed-method design. In this case, TM is not used instead of analysis "by hand", but becomes an addition: for instance, to identify representative texts or examine narrative structure prior to pursuing all the steps of a given qualitative method.

Finally, we conclude that the use of TM in textual analysis has direct ramifications for policy studies. The most evident benefit of TM is that it aids in the laborious work of textual analysis. As a result, existing policy concepts and theories can be applied to data sets of ambitious volume. In addition to the benefits of scale and scope, applying TM provides avenues for novel methodological approaches and combinations, which enable policy researchers to approach policy concepts, such as policy issues definitions and problem windows, in new ways. In particular, the characteristics of temporality and continuity in topics can offer useful avenues for interpretation. However, as topic interpretation depends

heavily on the case and context of research, the different synergies need to be applied carefully. We suggest that researchers think through the following issues when designing a study using TM:

1) How the technical assumptions of different TM variants can be aligned with the specifics of the phenomena of interest, and pre-processing and modelling parameters best set;

2) Which questions can reliably be answered by the corpus compiled and how corpus size and curation affects the potential results;

3) How to consider both parts of the output (word/topic and topic/document) proportion in assessing the phenomenon of interest;

4) How the interpretation of the topic output can be enhanced, validated, and criticised based on the document collection.

We expect this heuristic to help identify and address the critical points, thereby assisting researchers in the development of novel mixed-method designs that unlock the potential of TM in qualitative policy research without compromising methodological robustness.

 Bibliography

Atteveldt, Wouter van, Kasper Welbers, Carina Jacobi, and Rens Vliegenthart. 2014. "LDA Models Topics... But What Are 'Topics'?" *Big data in the Social Sciences workshop*.

Baumer, Eric P. S. et al. 2017. "Comparing Grounded Theory and Topic Modeling: Extreme Divergence or Unlikely Convergence?" *Journal of the Association for Information*

Science and Technology 68(6): 1397–1410.

Blei, D. M. 2012. Probabilistic Topic Models. Communications of the ACM, 55, 77-84.

Blei, David M., and John D. Lafferty. 2006. "Correlated Topic Models." *Advances in Neural Information Processing Systems 18*: 147–54.

Blei, David M, Andrew Y Ng, and Michael I Jordan. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 3: 993–1022.

Boussalis, Constantine, and Travis G. Coan. 2016. "Text-Mining the Signals of Climate Change Doubt." *Global Environmental Change* 36: 89–100.

Braun, V., and V. Clarke. 2006. "Using Thematic Analysis in Psychology." *Qualitative Research in Psychology* 3(May 2015): 77–101.

Bryman, A. 2006. "Integrating Quantitative and Qualitative Research: How Is It Done?" *Qualitative Research* 6(1): 97–113.

Carley, Kathleen. 1993. "Coding Choices for Textual Analysis: A Comparison of Content Analysis and Map Analysis." *Sociological Methodology* 23: 75.

Carley, Kathleen. 1994. "Extracting Culture through Textual Analysis." *Poetics* 22(4): 291–312.

Chandelier, Marie et al. 2018. "Content Analysis of Newspaper Coverage of Wolf Recolonization in France Using Structural Topic Modeling." *Biological Conservation* 220(January): 254–61.

Chang, Jonathan, Sean Gerrish, Chong Wang, and David M Blei. 2009. "Reading Tea

Leaves: How Humans Interpret Topic Models." *Advances in Neural Information Processing Systems 22*: 288--296.

Creswell, J W. 2003. "Research Design Qualitative Quantitative and Mixed Methods Approaches." *Research design Qualitative quantitative and mixed methods approaches*.

D 'Amato, Dalia et al. 2017. "Green, Circular, Bio Economy: A Comparative Analysis of Sustainability Concepts." *Journal of Cleaner Production* 168: 716–34.

Deerwester, Scott et al. 1990. "Indexing by Latent Semantic Analysis." *Journal of the American Society for Information Science*.

Dehghani, Morteza, Kenji Sagae, Sonya Sachdeva, and Jonathan Gratch. 2014. "Analyzing Political Rhetoric in Conservative and Liberal Weblogs Related to the Construction of the 'Ground Zero Mosque.'" *Journal of Information Technology & Politics* 11(1): 1–14.

Denny, M., & Spirling, A. (2018). Text Preprocessing For Unsupervised Learning: Why It Matters, When It Misleads, And What To Do About It. *Political Analysis*, 26(2), 168-189.

Denzin, Norman K, and Yvonna Lincoln. 2007. "Collecting and Interpreting Qualitative Materials. Third Edition." *SAGE Publications (CA)*.

DiMaggio, Paul. 2015. "Adapting Computational Text Analysis to Social Science (and Vice Versa)." *Big Data & Society* (December): 1–5.

DiMaggio, Paul, Manish Nag, and David Blei. 2013. "Exploiting Affinities between Topic Modeling and the Sociological Perspective on Culture: Application to Newspaper

Coverage of U.S. Government Arts Funding." *Poetics* 41(6): 570–606.

Ding, Chris, Tao Li, and Wei Peng. 2008. "On the Equivalence between Non-Negative

Matrix Factorization and Probabilistic Latent Semantic Indexing." *Computational*

Statistics and Data Analysis.

Entman, Robert M. 1993. "Framing: Toward Clarification of a Fractured Paradigm." *Journal*

of Communication 43(4): 51–58.

Fawcett, Paul, Michael Jensen, Hedda Ransan-Cooper, and Sonya Duus. 2018. "Explaining

the 'Ebb and Flow' of the Problem Stream: Frame Conflicts over the Future of Coal

Seam Gas ('Fracking') in Australia." *Journal of Public Policy*: 1–21.

Gamson, William A., and Andre Modigliani. 1989. "Media Discourse and Public Opinion on

Nuclear Power: A Constructionist Approach." *American Journal of Sociology* 95(1):

1–37.

Gilardi, F, CR Shipan, and B Wueest. 2017. "Policy Diffusion: The Issue-Definition Stage." :

1–55. http://fabriziogilardi.org/resources/papers/diffusion-policy-frames.pdf.

Glaser, Barney G, and Anselm L Strauss. 1967. "The Discovery of Grounded Theory."

International Journal of Qualitative Methods.

Goldstone, Andrew, and Ted Underwood. 2014. "The Quiet Transformations of Literary

Studies: What Thirteen Thousand Scholars Could Tell Us." *New Literary History: A*

Journal of Theory and Interpretation 45(3): 359–84.

Greene, Derek, and James P. Cross. 2017. "Exploring the Political Agenda of the European

Parliament Using a Dynamic Topic Modeling Approach." *Political Analysis* 25(1):

77–94.

Grimmer, Justin. 2010. "A Bayesian Hierarchical Topic Model for Political Texts: Measuring Expressed Agendas in Senate Press Releases." *Political Analysis* 18(1): 1–35.

Grimmer, Justin, and Brandon M. Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis* 21(3): 267–97.

Gritsenko, Daria. 2016. "Vodka on Ice? Unveiling Russian Media Perceptions of the Arctic." *Energy Research and Social Science* 16: 8–12.

Grubert, Emily, and Mark Algee-Hewitt. 2017. "Villainous or Valiant? Depictions of Oil and Coal in American Fiction and Nonfiction Narratives." *Energy Research and Social Science* 31: 100–110.

Grün, Bettina, and Kurt Hornik. 2011. "Topicmodels : An R Package for Fitting Topic Models." *Journal of Statistical Software*.

Hajer, Maarten A. 1995. *The Politics of Environmental Discourse*. Oxford University Press.

Hofmann, Thomas. 1999. "Probabilistic Latent Semantic Analysis." In *Uncertainty in Artificial Intelligence, UAI'99*, Stockholm.

Hsieh, Hsiu-Fang, and Sarah E Shannon. 2005. "Three Approaches to Qualitative Content Analysis." *Qualitative Health Research* 15(9): 1277–88.

Isoaho, Karoliina, Fanni Moilanen, and Arho Toikka. In press. "A Big Data View of the European Energy Union: Shifting from 'a Floating Signifier' to an Active Driver of

Decarbonisation?" *Politics and Governance* 7(1): X–X.

Jacobi, Carina, Wouter van Atteveldt, and Kasper Welbers. 2016. "Quantitative Analysis of Large Amounts of Journalistic Texts Using Topic Modelling." *Digital Journalism* 4(1): 89–106.

Jagarlamudi, Jagadeesh, Hal Daumé III, and Raghavendra Udupa. 2012. "Incorporating Lexical Priors into Topic Models." *Proc. of the 13th Conference of the European Chapter of the Association for Computational Linguistics, EACL '12*: 204–13.

Jaworska, Sylvia, and Anupam Nanda. 2016. "Doing Well by Talking Good: A Topic Modelling-Assisted Discourse Study of Corporate Social Responsibility." *Applied Linguistics* (Beattie 2014).

Keller, R. 2013. *Doing Discourse Research.* London, England: SAGE.

Kennedy, Mark Thomas. 2008. "Getting Counted: Markets, Media, and Reality." *American Sociological Review*.

Kim, Erin Hea-Jin et al. 2016. "Topic-Based Content and Sentiment Analysis of Ebola Virus on Twitter and in the News." *Journal of Information Science* 42(6): 763–81.

Klein, L. F., J. Eisenstein, I. Sun, and Jacko J. A. 2015. "Exploratory Thematic Analysis for Digitized Archival Collections." *Digital Scholarship in the Humanities* 30(suppl 1): i130–41.

Krippendorff, Klaus. 2004. 79 Education *Content Analysis: An Introduction to Its Methodology*.

Kuckartz, Udo, and Anne McWhertor. 2014. *Qualitative Text Analysis : A Guide to Methods, Practice & Using Software*. London: SAGE Publications Ltd

Lasswell, Harold Dwight, Daniel Lerner, and Ithiel de Sola Pool. 1952. 1 *The Comparative Study of Symbols: An Introduction*. Stanford University Press.

Light, Ryan, and Jeanine Cunningham. 2016. "Oracles of Peace: Topic Modeling, Cultural Opportunity, and the Nobel Peace Prize, 1902–2012 *." *Mobilization: An International Quarterly* 21(1): 43–64.

Lindgren, Simon. 2018. "A Ghost in the Machine : Tracing the Role of ' the Digital ' in Discursive Processes of Cybervictimisation." *Discourse & Communication*: 1–18.

Loewenstein, Jeffrey, William Ocasio, and Candace Jones. 2012. "Vocabularies and Vocabulary Structure: A New Approach Linking Categories, Practices, and Institutions." *Academy of Management Annals* 6(1): 41–86.

Lucas, Christopher et al. 2015. "Computer-Assisted Text Analysis for Comparative Politics." *Political Analysis* 23(02): 254–77.

Mcauliffe, J. D., and D. M. Blei. 2008. "Supervised Topic Models." *In Advances in neural information processing systems*: 121–28.

Mills, Kathy A. 2017. "What Are the Threats and Potentials of Big Data for Qualitative Research?" *Qualitative Research*: 146879411774346.

Mimno, David, and David M. Blei. 2011. "Bayesian Checking for Topic Models." In *EMNLP '11 Proceedings of the Conference on Empirical Methods in Natural Language*

Processing, , 227–37.

Mimno, David, and Andrew McCallum. 2008. "Topic Models Conditioned on Arbitrary Features with Dirichlet-Multinomial Regression." In *UAI'08 Proceedings of the Twenty-Fourth Conference on Uncertainty in Artificial Intelligence*, , 411–18.

Mishler, Elliot G. 1995. "Models of Narrative Analysis: A Typology." *Journal of Narrative and Life History* 5(2): 87–123.

Mohr, John W., and Petko Bogdanov. 2013. "Introduction-Topic Models: What They Are and Why They Matter." *Poetics* 41(6): 545–69.

Mohr, John W, and V Duquenne. 1997. "The Duality of Culture and Practice: Poverty Relief in New York City, 1888--1917." *Theory and Society* 26(2): 305–56.

Munksgaard, Rasmus, and Jakob Demant. 2016. "Mixing Politics and Crime – The Prevalence and Decline of Political Discourse on the Cryptomarket." *International Journal of Drug Policy* 35:

Murakami, Akira, Paul Thompson, Susan Hunston, and Dominik Vajn. 2017. "'What Is This Corpus about?': Using Topic Modelling to Explore a Specialised Corpus." *Corpora* 12(2): 243–77.

Nelson, Laura K. 2017. "Computational Grounded Theory." *Sociological Methods & Research*: 004912411772970.

Nigam, Amit, and William Ocasio. 2010. "Event Attention, Environmental Sensemaking, and Change in Institutional Logics: An Inductive Analysis of the Effects of Public Attention

to Clinton's Health Care Reform Initiative." *Organization Science*.

Nowlin, Matthew C. 2016. "Modeling Issue Definitions Using Quantitative Text Analysis." *Policy Studies Journal* 44(3): 309–31.

Pashakhin, Sergey. 2016. "Topic Modeling for Frame Analysis of News Media."

Quinn, Kevin M. et al. 2006. "An Automated Method of Topic-Coding Legislative Speech Over Time with Application to the 105th-108th U.S. Senate." *New York*: 61.

Ramage, Daniel, David Hall, Ramesh Nallapati, and Christopher D Manning. 2009. "Labeled LDA: A Supervised Topic Model for Credit Attribution in Multi-Labeled Corpora." *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1* (August): 248–56.

Riessman, Catherine Kohler. 2005. "Narrative Analysis." In *Narrative, Memory & Everyday Life*, , 1–8.

Roberts, Margaret E. et al. 2014. "Structural Topic Models for Open-Ended Survey Responses." *American Journal of Political Science* 58(4): 1064–82.

Rosen-Zvi, Michal et al. 2010. "Learning Author-Topic Models from Text Corpora." *ACM Transactions on Information Systems*.

Roth, Merrill C, and Eric P S Baumer. 2016. "It's the Definition, Stupid! Framing of Online Privacy in the Internet Governance ForumDebates." *Journal of Information Policy* 4(2014): 144–72.

Schofield, Alexandra, Måns Magnusson, and David Mimno. 2017. "Pulling Out the Stops :

Rethinking Stopword Removal for Topic Models." Pp. 432–36 in 15th Conference of the European Chapter of the Association for Computational Linguistics, vol. 2. Association for Computational Linguistics.

Schofield, Alexandra, Måns Magnusson, Laure Thompson, and David Mimno. 2017. Understanding Text Pre-Processing for Latent Dirichlet Allocation. Retrieved (http://www.cs.cornell.edu/~xanda/winlp2017.pdf).

Schön, Donald A., and Martin Rein. 1994. *Frame Reflection : Toward the Resolution of Intractable Policy Controversies*. New York: BasicBooks.

Seroussi, Yanir, Ingrid Zukerman, and Fabian Bohnert. 2014. "Authorship Attribution with Topic Models." *Computational Linguistics* 40(2): 269–310.

Silverman, David. 2016. *Qualitative Research.* London: SAGE Publications Ldt.

Tang Jian, Zhaoshi Meng, XuanLong Nguyen, Qiaozhu Mei, Ming Zhang. 2014. "Understanding the Limiting Factors of Topic Modeling via Posterior Contraction Analysis." *Proceedings of the 31 st International Conference on Machine Learning, Beijing, China: Volume 32.*

Titscher, S., M. Meyer, R. Wodak, and E. Vetter. 2000. *Methods of Text and Discourse Analysis: In Search of Meaning*. London, UK: SAGE.

Törnberg, Anton, and Petter Törnberg. 2016a. "Combining CDA and Topic Modeling: Analyzing Discursive Connections between Islamophobia and Anti-Feminism on an Online Forum." *Discourse & Society* 27(4): 401–22.

Törnberg, Anton, and Petter Törnberg. 2016b. "Muslims in Social Media Discourse:

Combining Topic Modeling and Critical Discourse Analysis." *Discourse, Context & Media* 13: 132–42.

Vaismoradi, Mojtaba, Hannele Turunen, and Terese Bondas. 2013. "Content Analysis and Thematic Analysis: Implications for Conducting a Qualitative Descriptive Study." *Nursing and Health Sciences* 15(3): 398–405.

APPENDIX 1

Method	Scopus search string	No. of articles 2008-2018
Sentiment analysis	ALL ("sentiment analysis") AND SRCTITLE (policy)) AND PUBYEAR > 2007 AND PUBYEAR < 2019	39
Word embeddings	ALL ("word embeddings" OR "word2vec") AND SRCTITLE (policy)) AND PUBYEAR > 2007 AND PUBYEAR < 2019	2
Supervised learning or supervised topic models	ALL ("supervised learning" OR "supervised topic model*") AND SRCTITLE (policy)) AND PUBYEAR > 2007 AND PUBYEAR < 2019	25
Supervised topic model	ALL ("supervised topic model*") AND SRCTITLE (policy)) AND PUBYEAR > 2007 AND PUBYEAR < 2019	2

LDA or topic model	ALL ("lda" *OR* "topic model*") AND SRCTITLE (policy)) AND PUBYEAR > 2007 AND PUBYEAR < 2019	78
Non-negative matrix factorisation	ALL ("non-negative matrix factorisation") AND SRCTITLE (policy)) AND PUBYEAR > 2007 AND PUBYEAR < 2019	4
Text clustering	ALL ("text clustering") AND SRCTITLE (policy)) AND PUBYEAR > 2007 AND PUBYEAR < 2019	5
Text analytics	ALL ("text analytics") AND SRCTITLE (policy)) AND PUBYEAR > 2007 AND PUBYEAR < 2019	5

This table lists the search strings used to trace the popularity of text mining methods. We acknowledge that the keyword list is not exhaustive. However, it is considered to give a comprehensive enough account of different methods to illustrate the use of topic modelling vis-à-vis other unsupervised or supervised methods.

APPENDIX 2

25 studies that have applied TM to conduct textual analysis in a qualitative setting

1. Atteveldt, W. van, Welbers, K., Jacobi, C. & Vliegenthart, R., 2014. LDA models topics... But what are 'topics'? Big data Soc. Sci. Work.

2. Bauer, P. C., Barberá, P., Ackermann, K. & Venetz, A., 2017. Is the Left-Right Scale a Valid Measure of Ideology?: Individual-Level Variation in Associations with "Left" and "Right" and Left-Right Self-Placement. Polit. Behav. 39, 553–583.

3. Baumer, E. P. S., Mimno, D., Guha, S., Quan, E. & Gay, G. K., 2017. Comparing grounded theory and topic modeling: Extreme divergence or unlikely convergence? J. Assoc. Inf. Sci. Technol. 68, 1397–1410.

4. Boussalis, C. & Coan, T. G., 2016. Text-mining the signals of climate change doubt. Glob. Environ. Chang. 36, 89–100.

5. Chandelier, M., Steuckardt, A., 2018. Mathevet, R., Diwersy, S. & Gimenez, O. Content analysis of newspaper coverage of wolf recolonization in France using structural topic modeling. Biol. Conserv. 220, 254–261.

6. Cody, E. M., Stephens, J. C., Bagrow, J. P., Dodds, P. S. & Danforth, C. M., 2017. Transitions in climate and energy discourse between Hurricanes Katrina and Sandy. J. Environ. Stud. Sci. 7, 87–10.

7. D 'Amato, D. et al. (2017)Green, Circular, Bio economy: a comparative analysis of sustainability concepts. Sustain. Sci. Day 2017 168, 716–734.

8. Fawcett, P., Jensen, M., Ransan-Cooper, H. & Duus, S., 2018. Explaining the 'ebb and flow' of the problem stream: frame conflicts over the future of coal seam gas ('fracking') in Australia. J. Public Policy 1–21.

9. Gill, A. J. et al., 2017. Insight workflow: Systematically combining human and computational methods to explore textual data. J. Assoc. Inf. Sci. Technol. 68, 1671–1686.

10. Goldstone, A. & Underwood, T., 2014. The Quiet Transformations of Literary Studies: What Thirteen Thousand Scholars Could Tell Us. New Lit. Hist. A J. Theory Interpret. 45, 359–384.

11. Grubert, E. & Algee-Hewitt, M., 2017. Villainous or valiant? Depictions of oil and coal in American fiction and nonfiction narratives. Energy Res. Soc. Sci. 31, 100–110.

12. Jacobi, C., van Atteveldt, W. & Welbers, K., 2016. Quantitative analysis of large amounts of journalistic texts using topic modelling. Digit. Journal. 4, 89–106.

13. Jaworska, S. & Nanda, A., 2016. Doing Well by Talking Good: A Topic Modelling-Assisted Discourse Study of Corporate Social Responsibility. Appl. Linguist.

14. Klebanov, B. B., Diermeier, D. & Beigman, E., 2008. Automatic Annotation of Semantic Fields for Political Science Research. J. Inf. Technol. Polit. 5, 95–120.

15. Klein, L. F., Eisenstein, J., Sun, I. & A., J. J., 2015. Exploratory Thematic Analysis for Digitized Archival Collections. Digit. Scholarsh. Humanit. 30, i130–i141.

16. Levy, K. E. C. & Franklin, M., 2014. Driving Regulation: Using Topic Models to Examine Political Contention in the U.S. Trucking Industry. Soc. Sci. Comput. Rev. 32, 182–194.

17. Light, R. & Cunningham, J., 2016. Oracles of Peace: Topic Modeling, Cultural Opportunity, and the Nobel Peace Prize, 1902–2012 *. Mobilization An Int. Q. 21, 43–64.

18. Lindgren, S., 2018. A ghost in the machine : Tracing the role of ' the digital ' in discursive processes of cybervictimisation. Discourse Commun. 1–18.

19. Lucas, C. et al., 2015. Computer-Assisted Text Analysis for Comparative Politics. Polit. Anal. 23, 254–277.

20. Munksgaard, R. & Demant, J., 2016. Mixing politics and crime – The prevalence and decline of political discourse on the cryptomarket. Int. J. Drug Policy 35, 77–83.

21. Nikolenko, S. I., Koltcov, S. & Koltsova, O., 2017. Topic modelling for qualitative studies. J. Inf. Sci. 43, 88–102.

22. Nowlin, M. C., 2016. Modeling Issue Definitions Using Quantitative Text Analysis. Policy Stud. J. 44, 309–331.

23. Pashakhin, S., 2016. Topic modeling for frame analysis of news media.

24. Törnberg, A. & Törnberg, P., 2016. Combining CDA and topic modeling: Analyzing discursive connections between Islamophobia and anti-feminism on an online forum. Discourse Soc. 27, 401–422.

25. Törnberg, A. & Törnberg, P., 2016. Muslims in social media discourse: Combining topic modeling and critical discourse analysis. Discourse, Context Media 13, 132–142.