Text as Data for Conflict Research: A Literature Survey

Seraphine F. Maerz and Cornelius Puschmann

Abstract Computer-aided text analysis (CATA) offers exciting new possibilities for conflict research that this contribution describes using a range of exemplary studies from a variety of disciplines including sociology, political science, communication studies, and computer science. The chapter synthesizes empirical research that investigates conflict in relation to text across different formats and genres. This includes both conflict as it is verbalized in the news media, in political speeches, and other public documents and conflict as it occurs in online spaces (social media platforms, forums) and that is largely confined to such spaces (e.g., flaming and trolling). Particular emphasis is placed on research that aims to find commonalities between online and offline conflict, and that systematically investigates the dynamics of group behavior. Both work using inductive computational procedures, such as topic modeling, and supervised machine learning approaches are assessed, as are more traditional forms of content analysis, such as dictionaries. Finally, cross-validation is highlighted as a crucial step in CATA, in order to make the method as useful as possible to scholars interested in enlisting text mining for conflict research.

Keywords Computer-aided text analysis · Text as data · Dictionary · Supervised and unsupervised machine learning · Topic models · Cross-validation

Seraphine F. Maerz and Cornelius Puschmann contributed equally to this chapter.

S. F. Maerz

V-Dem Institute, Department of Political Science, University of Gothenburg, Gothenburg, Sweden

C. Puschmann (✉)

Centre for Media, Communication and Information Research (ZeMKI), University of Bremen, Bremen, Germany

e-mail: cornelius.puschmann@uni-bremen.de

© The Author(s) 2020

E. Deutschmann et al. (eds.), *Computational Conflict Research*,
Computational Social Sciences, https://doi.org/10.1007/978-3-030-29333-8_3

1 Introduction

It is by now an old adage that the internet has transformed many areas of social life, from industry and politics to research and education. Computational techniques have benefited from this development through the rapid growth in open source software and cloud computing, both of which simplify research that utilizes computational approaches immensely, making them both simpler and less costly for social scientists to implement. However, there has also been a rapid growth in the availability of content to study—that is of text, images, and video—which is of relevance to social scientific inquiry. Focusing on text, such data range from administrative documents and digitized books to social media posts and online user comments. They also include traditional research data, such as open survey responses and interview transcripts, which may be scrutinized with computational techniques.

The field of computer-aided text analysis (CATA) subsumes methods used to study such data. The goal of this chapter is to provide an overview of computational methods and techniques related to the area of (semi)automated content analysis and text mining, with emphasis on the application of such approaches to conflict research. We describe three central areas of CATA in order of their respective age: techniques relying on dictionaries and simple word counting, supervised machine learning (SML), and unsupervised machine learning (UML). While doing this, we provide a survey of published studies from a variety of fields that implement CATA techniques to study conflict. We then proceed to address issues of validation, a particularly important area of CATA.

Throughout the chapter, we offer a host of examples of how the application of CATA may advance conflict research. Our working definition of conflict in this chapter is twofold: we cite studies using CATA to study violent conflict on a regional or national level, usually by means of relating textual data that applies to a particular actor (for example, a country) to some indicator of violence. Such studies aim to uncover hidden relationships between issues, frames, and rhetoric on the one side and violent conflict on the other. The second branch of studies that we cite are those where conflict is non-violent but there is a considerable aggressive potential, for example, in online hate speech campaigns, cyber mobbing, and social media flame wars. Such studies are as diverse as their respective objects, but a commonality is that because there is usually ample data to document the conflict, CATA may be used to draw a precise picture of the actors, issues, and temporal dynamics. By presenting both branches of physical and virtual conflict research side by side, we do not imply that one follows from the other, but rather that the same approaches may be useful in studying both.

The techniques that we describe can be seen as existing on a continuum, from approaches that are more deductive in nature and presuppose very detailed domain knowledge and precise research questions/hypotheses, such as dictionary analysis, to (more) inductive methods such as unsupervised learning that are more suitable for exploration (cf. Table 1). The latter methods also tend to be more

Table 1 CATA methods for conflict research, adjusted from (Boumans and Trilling, 2016, p. 10)

	Dictionary	Supervised (SML)	Unsupervised (UML)
Typical research contexts/material in conflict studies	Sentiment analysis of documents from opposing parties or extremist groups, or time series of sentiment fluctuation	Relatively homogeneous, numerous texts, e.g. from newspapers; sentiment classification of sentences or paragraphs	Large amounts of unexplored material, e.g. from field research, official documents, social media
Common statistical procedures	Counting of word frequencies, string comparisons	Support vector machines, Naive Bayes, Neural Networks	(Structural) topic modeling, latent Dirichlet allocation
Reasoning	<div><div>Deductive</div><div>Inductive</div><div></div></div>		

computationally resource-intensive than the former, though this will only really be felt when truly large volumes of data are analyzed on a regular desktop or laptop computer, and they tend to be more opaque and subject to interpretation than simple dictionary techniques which have been in use for decades. However, this is not truly a dichotomy as in typical CATA workflows multiple methods are often combined in different stages of the research. This can both serve the purpose of developing one resource based on the output of another (for example, developing a topical dictionary based on the results of unsupervised machine learning) or on the validation of a particular technique with another.¹

Similar overviews of CATA for other fields have been provided before, for example, in political science, communication studies, and sociology (Boumans and Trilling, 2016; Grimmer and Stewart, 2013; DiMaggio, 2015). We aim to extend this body of work with an overview of research in conflict research that will be useful to computational social scientists aiming to use CATA in their work.

2 Dictionary Approaches for Conflict Research

Dictionary methods are among the oldest techniques employed in text mining and automated content analysis in the social sciences (Stone et al., 1966) and are popular in part due to their simplicity and transparency when compared with more recent methods (Grimmer and Stewart, 2013). In fact, dictionary approaches are both comparatively easy to interpret and computationally cheap, making them popular across a wide range of fields and research subjects. Dictionary approaches rely on

¹For a comprehensive, hands-on introduction to CATA with Python for social scientists see Trilling (2018), for a similar introduction with R (in German) see Puschmann (2018).

the frequency of specific words (those contained in the dictionary) to assign each document in a corpus to a category. For example, a list of words describing violent conflict may be used to operationalize the topic, allowing the researcher to gauge the level of debate of this issue over time or by actor, or such a list may be used to identify potentially relevant material in a larger corpus.² Specialized topical or psychological dictionaries as they are used within the social sciences should not be confused with linguistic techniques such as part of speech-tagging, syntactic parsing, or named entity recognition (NER), which also allow the reduction of words to aggregate categories (nouns, sentence subjects, place names, etc.), but are usually intended to describe linguistic form rather than social or communicative function.

In some implementations, membership in a dictionary category is proportional to the number of words occurring in the text that belong to that category, while in others a winner-takes-all approach is used in which the document is assigned to the single category with the largest number of matching terms. The difference between the two styles is the weighting applied to the document feature matrix which contains the dictionary terms and the texts in which they occur, which is usually conducted after the words are counted (Grimmer and Stewart, 2013, p. 274).

The increasingly popular method of sentiment analysis—also called opinion mining in computer science and computational linguistics—is in many cases a simple variant of dictionary analysis in which the dictionary terms belong to one of two categories, positive or negative sentiment (although sentiment dictionaries with three or more types of sentiment also exist). Sentiment dictionaries exist in many variants across languages,³ text types, and applications and are often quite comprehensive when compared with specialized topical lexicons. In the case of binary classification (which applies to many forms of sentiment analysis), the logarithm of the ratio of positive to negative words is often used to calculate a weighted composite score (Proksch et al., 2019).

Other dictionaries also exist in a wide variety of shapes and formats, and for a large number of different applications (Albaugh et al., 2013; Burden and Sanberg, 2003; Kellstedt, 2000; Laver and Garry, 2000; Young and Soroka, 2012). These include policy areas, moral foundations and justifications, illiberal rhetoric as well as place and person names, and other strongly standardized language use. Such off-the-shelf dictionaries provide a level of validity by being widely used and (in some cases) even being able to assign material in different languages to similar categories by having corresponding word lists for each category (Bradley and Lang, 1999; Hart, 2000; Pennebaker et al., 2001). Dictionaries can also be created through a variety of techniques, including using manually labeled data from which the most distinctive terms can be extracted.

A key strength of dictionary approaches (and of both supervised and unsupervised learning) is their ability to reduce complexity by turning words into category

²See Payson Conflict Study Group (2001) for such a list.

³See, for example, Proksch et al. (2019) for a multilingual sentiment analysis based on automatically translated dictionaries.

distributions. The basis of this approach is what is known as the bag-of-words philosophy of text analysis which turns a sequence of words and sentences into an undifferentiated “bag” which records only the frequency of each word within each text, but no information on where in a text a particular word occurs (Lucas et al., 2015, p. 257). Oftentimes this is not a hindrance, as in most quantitative research designs scholars will be interested primarily in distilling some aggregate meaning from their data, rather than retaining its full complexity. This decision entails a number of trade-offs, however, from a loss of structure and meaning that occurs when a text is pre-processed and cleaned to the alignment of the dictionary categories with the specific meaning of the material under study. The loss of syntactic information and argument structure is also an important limitation in bag-of-word approaches, which are often used in dictionary analysis (though dictionaries of n-grams are both technically possible and in widespread use).

Dictionaries have long played an important role in conflict research. Baden and Tenenboim-Weinblatt (2018) rely on a custom-built cross-linguistic dictionary of more than 3700 unique concepts, including actors, places, events, and activities which they use to study the media coverage of six current violent conflicts in domestic and international media over time. While compiling such a dictionary is burdensome, machine translation can be used to turn a mono-linguistic dictionary into one covering corresponding concepts across languages. Person and place names, specific events, and actions can all be captured by such a dictionary with relative accuracy, underlining why such a simple approach can be extremely effective (Baden and Tenenboim-Weinblatt, 2018), though translation always needs careful validation from experts. A broadly similar approach is used by Brintzenhoff (2011) who relies on a proprietary software to identify instances of violent conflict. There are also examples of studies that rely on data mining to generate dictionaries or resources similar to them. Montiel et al. (2014) present an analysis of the national news coverage on the Scarborough Shoal conflict between the Philippines and China relying on RapidMiner, a commercial machine learning software suite. A principal component analysis differentiates specific issues that are specific to Filipino and Chinese news sources from each other.

Dictionaries are also used to study conflict in virtual environments. Ben-David and Matamoros-Fernández (2016) rely on simple word frequencies in their study of hate speech on the Facebook pages of extreme-right political parties in Spain. After cleaning the data and removing stopwords, they group posts according to broad thematic categories and then extract those terms most frequently within each group, yielding category descriptions of different groups of immigrants and other “enemies.” This approach is then combined with an analysis of hyperlinks and visual data. Broadly similar, Cohen et al. (2014) suggest identifying specific categories of radicalization as they manifest in “lone wolf” terror subjects through a combination of ontologies such as WordNet (Miller, 1995) and dictionaries such as LIWC (Tausczik and Pennebaker, 2010). While their overview is rather general, it points to the potential of composite solutions for linking behavior and language use.

As Grimmer and Stewart (2013) note, problems occur when dictionaries from one area are applied in another domain, leading to potentially serious errors when the problem is not caught. The authors cite the example given by Loughran and McDonald (2010) in which corporate earnings reports that mention terms such as “cancer” or “crude” (oil) are assigned negative sentiment scores, even when health care or energy firms mention these terms in an entirely positive context. This problem may seem entirely unsurprising, but particular assumptions about the nature of language (and in many cases writing) lead to the belief that a specialized dictionary that is appropriate in one domain will also produce valid results in another. As the example shows, even something as presumably universal as sentiment is a case in point: a dictionary that is suitable for capturing the opinion of a consumer about a product in a review on a shopping site will not produce equally valid results when applied to political speeches or newspaper articles, because (1) in them the same words may express different meanings, (2) such texts are presumably much more neutral in tone to begin with, (3) such texts do not necessarily express the opinion of their author, but institutional viewpoints, and (4) such texts report on or respond to the opinions of others.

Dictionaries should always be validated using the data to which the dictionary is to be applied, in other words it should not be presumed that the dictionary will produce accurate results if it is applied to a domain that is in any way different from the one for which it was developed. This applies equally to off-the-shelf and self-made dictionaries. Systematically validating dictionary results, for example, by means of traditional content analysis, is one common pathway to overcoming these problems.

3 Supervised Methods

Supervised machine learning (SML) represents a significant step away from the useful but also quite limited methods described in the previous section, towards more advanced techniques that draw on innovations made in the fields of computer science and computational linguistics over the past 30 years. This does not mean that such techniques are generally superior to dictionary approaches or other methods that rely on word counting, but that they utilize the extremely patterned nature of word distributions. In particular, supervised machine learning is able to connect feature distribution patterns with human judgment by letting human coders categorize textual material (sentences, paragraphs, or short texts) according to specific inferential criteria and then asking an algorithm to make a prediction of the category of a given piece of text based on its features. Once a classifier has been trained to a satisfactory level of accuracy, it can be used to classify unknown material. The algorithm thus learns from human decisions, allowing for the identification of patterns that humans are able to discern, but that are otherwise not obvious with methods relying purely on words and word distribution patterns.

The perhaps most typical research design consists of a set of labeled texts (alternatively paragraphs, sentences, social media messages, rarely more complex syntactic structures) from which it is possible to derive feature distributions, typically words (alternatively n-grams, part of speech information, syntactic structures, emojis). First, the data is split into a training and a test data set. An algorithm then learns the relation of the label to the distribution of the features from the training data set and then applies what has been learned to the test data set. This produces a set of metrics which allow to evaluate the classifier's performance. If the quality of the automated coding is deemed as satisfactory (i.e., similar to or better than human annotation) in terms of its precision and recall, the classifier can be applied to new, previously uncoded material. There are three major uses to this basic technique, including the validation of a traditional content analysis, the automated annotation of unknown material, and the discovery of structural relationships between external variables that prove to be reliable predictors for language use (Puschmann, 2018).

The applications of SML to conflict research and to social science more broadly are manifold. In a traditional content analysis, achieving a high inter-coder reliability is usually a key aim, because it signals that a high degree of inter-subjectivity is feasible when multiple humans judge the same text by a previously agreed set of criteria. In this approach, the machine learning algorithm in effect becomes an additional "algorithmic coder" (Zamith and Lewis, 2015) that can be evaluated along similar lines as a human would be. Crucially, in such an approach the algorithm aims to predict the—presumably perfect—consensus judgment of human coders that is treated as "ground truth." Social scientists who rely on content analysis know that content categories are virtually never entirely uncontroversial. Since obviously humans disagree with one another, there is a risk of "garbage in, garbage out" when training the classifier on badly annotated material. Thus, the quality of the annotation and the linguistic closeness of the relation between content and code is the key, and the notion of "ground truth" should be treated with care.

This is usually not an issue when what is being predicted is the topic or theme of a text. For example, Scharkow (2013) relies on SML to gauge the reliability of machine classification in direct comparison to human coders, comparing the topics assigned to 933 articles from a range of German news sources. He finds automated classification to yield very good results for certain categories (e.g., *sports*) and poor results for others (e.g., *controversy* and *crime*), with implications for conflict research. As the author points out, even for categories where the classification results are less reliable, the application of SML yields important findings on the quality of manual content analyses. Similarly, van Atteveldt et al. (2008) are able to predict different attributes and concepts in a manually annotated corpus of Dutch newspaper texts using a range of lexical and syntactic features for their prediction. In both bases, the SML approach yields good results because the annotation is of high quality and the categories that are being predicted are strongly content-bound, rather than interpretative.

While frequently the categories coded for are determined through content analysis and relatively closely bound to the text itself (themes, issues, frames,

arguments), or can be related to social or legal norms (e.g., hate speech), it is worth noting that any relevant metadata may be used as the label that the classifier aims to make a prediction on. For example, Kananovich (2018) trains a classifier on a manually labeled data set of frames in international news reports that mention taxes, and tests two hypotheses related to the prevalence of certain frames in countries with particular political systems.

Burscher and colleagues have shown that supervised machine learning can be used to code frames in Dutch news articles and reliably discern policy issues (Burscher et al., 2014, 2015). Sentiment analysis using SML has also been applied, with results considerably better than those of approaches that are purely based on the application of lexicons (González-Bailón and Paltoglou, 2015).

Burnap and Williams (2015) train a sophisticated supervised machine learning text classifier that distinguishes between hateful and/or antagonistic responses with a focus on race, ethnicity, or religion; and more general responses. Classification features were derived from the content of each tweet, including grammatical dependencies between words to recognize “othering” phrases, incitement to respond with antagonistic action, and claims of well-founded or justified discrimination against social groups. The results of the classifier draw on a combination of probabilistic, rule-based, and spatial classifiers with a voted ensemble meta-classifier.

Social media data can also be productively combined with demographic and geo-spatial data to make predictions on issues such as political leanings. For example, Bastos and Mercea (2018) fit a model that is able to predict support for the Brexit referendum in the UK based on the combination of geo-localized tweets and sociodemographic data.

Though manual classification is the norm, in some cases, a combination of unsupervised and supervised machine learning may yield good results. Boecking et al. (2015) study domestic events in Egypt over a 4-year period, effectively using the metadata and background knowledge of events from 1.3 million tweets to train a classifier.

Other approaches that connect manual content analysis with supervised machine learning that are presently still underutilized in the social sciences include argumentation mining. For example, Bosc et al. (2016) provide an overview of argument identification and classification using a number of different classifiers applied to a range of manually annotated Twitter data sets. Using a broader range of features in particular appears to increase the performance of SML techniques markedly.

4 Topic Modeling as Unsupervised Method in Conflict Research

The main difference between supervised and unsupervised text as data methods is that unsupervised techniques do not require a conceptual structure that has been defined beforehand. As explained above, dictionary applications and supervised

techniques are *deductive* approaches which rely either on a theoretically informed collection of key terms or a manually coded sample of documents to specify what is conceptually interesting about the material before applying a statistical model to extend the insights to a larger population of texts. In contrast to this, unsupervised methods work *inductively*: without predefined classification schemes and by using relatively few modeling assumptions, such algorithm-based techniques shift human efforts to the end of the analysis and help researchers to discover latent features of the texts (Lucas et al., 2015, p. 260, Grimmer and Stewart, 2013).

Unsupervised text as data techniques are useful for conflict research—especially for understudied areas and previously unknown primary sources or the many rapidly growing digitized resources—because they have the potential to disclose underlying clusters and structures in large amounts of texts. Such new insights can either complement and refine existing theories or contribute to new theory-building processes about the causes and consequences of conflict.

While there are several variations of unsupervised methods,⁴ our literature survey shows that topic modeling is the most frequently used technique in conflict research. Common to topic modeling is that *topics* are defined as probability distributions over words and that each document in a corpus is seen as a mixture of these topics (Chang et al., 2009; Grimmer and Stewart, 2013; Roberts et al., 2014). The first and still widely applied topic model is the so-called LDA—latent Dirichlet allocation (Blei et al., 2003; Grimmer and Stewart, 2013). Recently, the Structural Topic Model (STM) has been proposed as an innovative and increasingly used alternative to the LDA (Roberts et al., 2014; Lucas et al., 2015; Roberts et al., 2016). Whereas the LDA algorithm assumes that topical prevalence (the frequency with which a topic is discussed) and topical content (the words used to discuss a topic) are constant across all documents, the STM allows to incorporate covariates into the algorithm which can illustrate potential variation in this regard (Roberts et al., 2014, p. 4).

Typically, the workflow⁵ of topic modeling starts with a thorough cleaning of the text corpus, as commonly done for quantitative bag-of-words analyses which transform texts into data. Depending on the research focus, such automated preprocessing includes lowercasing of all letters, erasing of uninformative non-letter characters and numbers, stopword removal, stemming, and possibly also the removal of infrequently used terms. Text cleaning procedures can have significant and unexpected effects on the results of unsupervised analyses which is why Denny and Spirling (2018) recommend “reasonable” preprocessing decisions and suggest a new technique to test their potential effects.⁶ Subsequently, researchers must make some model specifications such as determining the number of topics (K) to

⁴Grimmer and Stewart (2013) provide a useful overview in this regard.

⁵While some researchers perform topic modeling in *Python* (2018), the detailed vignettes as well as online support and tutorials of packages in *R* (2018) such as *quanteda* (Benoit et al., 2018) or *stm* (Roberts et al., 2018) make these tools easily accessible.

⁶For detailed explanations of how to apply these tests see Denny and Spirling (2018).

be inferred from the corpus and—in case of the STM—the choice of covariates. Through Bayesian learning, the model then discriminates between the different topics in each document. Concretely this means, for example, that based on updated word probabilities, the algorithm would group terms such as “god,” “faith,” “holy,” “spiritual,” and “church” to one topic in a document, while the same document could also contain words such as “bloody,” “violent,” “death,” “crime,” and “victim” constituting a second topic. Lastly, it is the researchers’ task to adequately label and interpret such topics and make more general inferences.

Topic modeling is a new methodological trend in conflict research—the recent growth in studies which apply such methods point to the great potential these innovative approaches have in this area. Examples cover a broad range of issues: Stewart and Zhukov (2009), for instance, analyze nearly 8000 public statements by political and military elites in Russia between 1998 and 2008 to assess the country’s public debate over the use of force as an instrument of foreign and defense policy. The LDA analysis of Bonilla and Grimmer (2013) focuses rather on how external threats of using force and committing a terrorist attack influence the themes of major US media and the public’s policy preferences at large. Other studies applying the LDA algorithm scrutinize patterns of speaking about Muslims and Islam in a large Swedish Internet forum (Törnberg and Törnberg, 2016) or generally look into how controversial topics such as nuclear technology are discussed in journalistic texts (Jacobi et al., 2016). While Fawcett et al. (2018) analyze the dynamics in the heated public debate on “fracking” in Australia as another example of non-violent conflict, Miller (2013) shows that topic modeling can be also valuable to study historical primary sources on violent crimes and unrest in Qing China (1722–1911).

One central and rather broad contribution to conflict research is the study of Mueller and Rauh (2018). Based on LDA topic modeling, they propose a new methodology to predict the timing of armed conflict by systematically analyzing changing themes in large amounts of English-speaking newspaper texts (articles from 1975 to 2015, reporting on 185 countries). The added value of using unsupervised text-mining techniques here is that the explored within-country variation of topics over time help to understand *when* a country is at risk to experience violent outbreaks, independent of whether the country had experienced conflicts in the past. This is truly innovative because earlier studies could merely predict a general, not time-specific risk in only those countries where conflict had appeared before. Mueller and Rauh (2018, p. 359) combine their unsupervised model with panel regressions to illustrate that (not) reporting on particular topics increases the likelihood of an upcoming conflict. They show, for example, that the reference to judicial procedures significantly decreases before conflicts arise.

Other recent conflict analyses apply the newly proposed STM model of Roberts et al. (2014, 2016, 2018). As explained above, the difference between the LDA and STM algorithm is that the latter allows to include document-level metadata. Lucas et al. (2015), for example, specify in their model on Islamic *fatwas* whether clerics are Jihadists or not. Based on this, they illustrate crucial topical differences between both groups—thus, Jihadists mostly talk about “Fighting” and “Excommunication”

while non-Jihadists rather use topics such as “Prayer” and “Ramadan” (Lucas et al., 2015, p. 265). Terman (2017) uses STM to scrutinize Islamophobia and portrayals of Muslim women in US news media. Her findings of analyzing a 35-year coverage of journalistic texts in the *New York Times* and *Washington Post* (1980–2014) on women in non-US countries show that stories about Muslim women mostly address the violation of women’s rights and gender inequality while stories about non-Muslim women emphasize other topics. Further research on conflicts which also make use of STM include Bagozzi and Berliner’s (2018) analysis of crucial variations over time concerning topic preferences in human rights monitoring or Mishler et al. (2015) test of detecting events based on systematically analyzing Ukrainian and Russian social media.

Validating model specifications and particularly the labeling and interpretation of topics as model output is an absolutely crucial part of any unsupervised text analysis. As Grimmer and Stewart (2013) point out, such post-fit validation can be extensive. However, systematic validation procedures and standardized robustness tests for unsupervised methods are still pending. Frequently, applications of topic models in conflict research and other fields of study exhibit two shortcomings in this regard: First, the model specification of determining the number of topics (K) is not sufficiently justified. Second, the labeling and interpretation of topics seem arbitrary due to lack of information about this process.

The selection of an appropriate number of topics (K) is an important moment in topic modeling: too few topics result in overly broad and unspecific categories, too many topics tend to over-cluster the corpus in marginal and highly similar topics (Greene et al., 2014, p. 81). The general aim in this regard is to find the number of K that yields the most interpretable topics. While there are methods and algorithms to automatically select the number of topics (Lee and Mimno, 2014; Roberts et al., 2018), Chang et al. (2009) show that the statistically best-fitting model is usually not the model which provides substantively relevant and interpretable topics. To reach this goal, we recommend to conduct systematic comparisons of model outcomes with different K s, similar to Bagozzi and Berliner (2018), Jacobi et al. (2016), Mueller and Rauh (2018), Lucas et al. (2015). Visualizations of such robustness tests such as in Maerz and Schneider (2019) further increase the transparency concerning the decision-making process of determining K .

A valid process of labeling and interpreting topics as model outcome includes a thorough analysis of the word profiles for each topic. While computational tools can efficiently support such examinations, one should keep in mind that this is a genuinely interpretative and rather time-consuming act which needs to be documented in a comprehensible manner. The *R* package *stm* offers several functions to visualize and better understand the discursive contexts of topics (Roberts et al., 2018). This includes the compilation of detailed word lists with most frequent and/or exclusive terms per topic (*tableTopics*), the qualitative check of most typical texts for each topic (*findThoughts*), or estimating the relationship between metadata and topics to better understand the context and interrelation of the topics at large (*estimateEffect*). In addition, Schwemmer’s (2018) application

stminsights provides interactive visualization tools for STM outcomes to facilitate a straightforward validation. In the following section, we make several suggestions of how to further strengthening the validity of automated content analysis in conflict research by combining topic modeling with other text-mining techniques and quantitative or qualitative methods.

5 Techniques of Cross-Validation

In their groundbreaking article on automated content analysis of political texts, Grimmer and Stewart (2013, p. 269) suggest four principles of this method: (1) *While all quantitative models of language are wrong, some are indeed useful.* (2) *Automated text analysis can augment, but not replace thorough reading of texts.* (3) *There is no universally best method for quantitative text analysis.* (4) *Validate, validate, validate.* It is particularly the latter point which we would like to emphasize in this section. Automated text analysis has the potential to significantly reduce the costs and time needed for analyzing a large amount of texts in conflict research—yet, such methods should never be used blindly and without meticulous validation procedures that illustrate the credibility of the output.

As we have argued above, the validation of dictionary approaches and supervised techniques needs to show that such methods can replicate human coding in a reliable manner (Grimmer and Stewart, 2013, p. 270). For unsupervised methods, it is important to justify and explain model specifications and demonstrate that the model output is conceptually meaningful. Beside these necessary steps for each method individually, we recommend to combine dictionary approaches and supervised as well as unsupervised techniques as efficient tools for cross-validation. In agreement with Grimmer and Stewart (2013, p. 281) we hold that these different techniques are highly complementary and suggest two strategies of designing such multi-method validations. The first procedure of cross-validation is rather inductive and particularly suitable for exploring new theoretical relations and conceptual structures in large amounts of hitherto broadly unknown texts. This technique is similar to what Nelson (2017) describes as “computational grounded theory.” Figure 1 provides a simplified illustration of this process, which we refer to as the inductive cycle of cross-validation. The starting point of this framework is topic modeling because it allows for an inductive computational exploration of the texts. Nelson (2017) calls this the pattern detection step, which subsequently facilitates the formulation of new theories. Based on this theory-building process, a targeted dictionary or coding scheme is conceptualized. The outcome of applying this newly developed dictionary or coding scheme can illustrate that the results of the preceding topic modeling are indeed conceptually valid and—to a certain degree—comparable to measures from supervised models (Grimmer and Stewart, 2013, p. 271). Furthermore, such supplementary supervised analyses are more focused

Fig. 1 The inductive cycle of cross-validation

and helped to illuminate specific aspects of the texts which are theoretically more interesting than the broad outcome of the explorative topic modeling.

The rich and original material gained during ethnographic field research is one example from conflict studies for which the inductive cycle would be a suitable approach. After having conducted open-ended surveys in a country torn by ethnic conflicts, for instance, one is confronted with huge amounts of unique texts which ask to be analyzed. Topic modeling is a fruitful start in this regard (Roberts et al., 2014), followed by a more fine-grained and theory-guided dictionary analysis or supervised learning. Overall, the suggested framework allows for a thorough cross-validation of the different analytic steps and is a comprehensive way of computationally accessing new information—in this example about the nature of ethnic conflicts.

The second procedure of cross-validation is a deductive approach that implies that the researcher has an existing theoretical framework in mind when developing a dictionary or coding scheme for supervised learning. Alternatively, one could also apply an already established dictionary to a corpus of texts for which this application is theoretically and substantially justified (yet, see Sect. 2 regarding the risks of blindly adopting dictionaries for diverging fields of inquiry). As illustrated in Fig. 2, this first step is followed by a topic model applied to the same corpus of texts to additionally explore hidden features in the material that might be of theoretical interest but are not yet captured by the dictionary or coding scheme. The outcome of the topic modeling—typically a report of top terms appearing in K topics—has then the potential to validate but also significantly complement and refine the existing dictionary or coding scheme, leading to more solid results.

Fig. 2 The deductive cycle of cross-validation

The analysis of propaganda magazines or online material published by a newly emerging Islamist terrorist group is one example from conflict research that could be adequately analyzed with the described deductive framework. Making use of existing theories about Islamist communication strategies or applying an already established dictionary that was developed to analyze Islamist rhetoric seems adequate to scrutinize the content of such texts in a first step. However, since the assumed terrorist group would be a new formation in the field of Islamist fundamentalism, the additional application of topic modeling could disclose so far unknown aspects about this group or the language of terrorists in general. This, in turn, contributes to further improving the existing dictionary or coding scheme and, overall, enables a more valid analysis.

Existing empirical analyses from related fields of research that apply a similar validation cycle include the study of the language of autocrats by Maerz (2019) or the analysis of illiberalism in the speeches of political leaders by Maerz and Schneider (2019). The latter further expand the validity tests to qualitative checks and network analysis to handle their particularly heterogeneous material. While we have focused here solely on a fruitful combination of various text as data techniques, the inclusion of other qualitative and quantitative methods and visualization techniques is another option to further test and illustrate the validity of the results.

6 Conclusion

In this chapter, we discussed several CATA methods used in conflict research as new techniques to handle growing amounts of written on- and offline resources. Table 2 compares the performance of the different approaches which we have described in the preceding sections. The first technique—dictionary applications—is rather straightforward and comparatively easy to apply once a theory-guided selection of keywords has been defined. For conflict researchers interested in text mining methods, this first approach might be particularly suitable if material was collected from a research field that is already widely covered by established theories. The dictionary analysis could help, for example, to further refine those theories. Yet, as Table 2 specifies, one disadvantage of dictionary applications is that it can be very challenging to justify why a certain selection of terms is more suitable than alternative word lists. Such procedures typically imply extensive qualitative procedures to illustrate the validity of the dictionary.

The second approach we discussed is supervised machine learning. Supervised text mining is a more sophisticated approach than dictionary applications because it is not limited to a fixed list of keywords. Instead, these semi-automated methods make use of algorithms which *learn* how to apply the categories of a manually coded training set to larger amounts of texts. One downside of supervised learning is that the manual coding of the training set can be highly work-intensive. This is why we recommend this method for conflict researchers who are either experienced in the manual coding of texts or have sufficient capacities to handle this first and laborious step of the analysis.

Lastly, we reviewed topic modeling as the most current unsupervised method applied in conflict research. Topic modeling is particularly suitable for sizable amounts of new texts that cannot be manually screened since these methods help to explore the underlying structure and topics of the hitherto unknown texts. While this inductive detection of topics is fully automated, the definition of model specifications and interpretation of the model outcome require high human efforts and transparency to ensure valid and non-arbitrary inferences (cf. Table 2).

Table 2 Comparing the performance of different CATA methods

	Dictionary	Supervised (SML)	Unsupervised (UML)
Advantages	Relatively easy to apply, no comprehensive cleaning of corpus needed	Not limited to fixed keywords, extends manually coded training set to large amounts of texts	Reveals latent features in (hitherto unknown) texts, analysis is fully automated
Disadvantages	Validation of dictionary challenging, re-usage of dictionary in different contexts problematic	Laborious manual coding of training set required	Model specifications and interpretation of model outcome requires high human efforts

As one recommendation for future text mining projects in conflict research, we highlighted validation as a crucial element of all text as data methods. Ideally, tests for validity evaluate model performance and compare the output of the model to the results of hand coding, illustrating that the automated analysis closely replicates the human-coded outcome. However, applying such procedures can be costly and difficult to implement in many settings. This is why we additionally suggested two cycles of combining dictionary approaches, supervised methods and unsupervised techniques to effectively cross-validate the outcome of these applications.

Apart from extensive validation procedures, we believe that transparency in terms of methodological decisions and steps, accessibility to data and replication files as well as open access publications are critical to advance computational methods in conflict research and beyond. While researchers have started to follow these practices in providing online appendices on methodological details and robustness tests and making their replication files publicly available on dataverses,⁷ there is still a large number of studies that are rather nebulous about these things, further enforcing the much-discussed replication crisis in the social sciences. Text as data approaches are currently experiencing a hype—yet, while plenty of innovative tools and techniques are being developed, there is the need for platforms and digital hubs that bundle the newly gained knowledge and make it accessible to a broader community of researchers.⁸ Such new policies of data sharing and digital cooperation pave the way for a more networked and progressive computational methodology in the social sciences.

Appendix

See Table 3.

⁷For example, <https://dataverse.harvard.edu/>.

⁸First steps into this direction are initiated by research institutes such as the Social Media and Political Participation Lab (<https://smappnyu.org/>), the MediaCloud (<https://mediacloud.org/>), the Berkman Klein Center for Internet and Society at Harvard (<https://cyber.harvard.edu/>), the Oxford Internet Institute (<https://www.oii.ox.ac.uk/>) or the newly founded Digital Democracy Lab (<https://digdemlab.github.io/>).

Table 3 This overview is not an exhaustive list but rather a selection of text mining examples in the field of conflict research

Author(s)	Conflict	Data	Text mining technique	Software ^a	Validation ^a
Baden and Tennenboim-Weinblatt (2018)	Six selected armed conflicts	896,480 news texts from 66 news outlets	Dictionary based on cross-linguistic semantic concepts	<i>R</i>	–
Bagozzi and Berliner (2018)	Human Rights monitoring	6,298 State Department Country Reports on Human Rights practices (1977–2012)	STM topic modeling	<i>R</i>	Alternative K_s^b
Bastos and Mercea (2018)	Brexit	8,821,116 tweets, 10,000 of which were hand-coded	Supervised learning	<i>R</i>	Combined with hand-coding
Burnap and Williams (2015)	Woolwich murder	450,000 tweets, 2,000 of which were hand-coded	Supervised learning	<i>Java</i> (Weka)	Combined with crowdsourced hand-coding
Fawcett et al. (2018)	Frame conflicts over the future of ‘fracking’ in Australia	Relevant newspaper articles	LDA topic modeling	–	Combined with hand-coding
Greene et al. (2018)	Human rights violations	US State Department Human Rights Reports (1978–2010)	Supervised learning algorithms	<i>R</i>	Cross-validation with different models
Macnair and Frank (2018a)	Islamic State	IS propaganda magazines <i>Dabiq</i> and <i>Rumiyah</i> (25 issues)	Sentiment analysis (over time)	<i>SentiStrength</i>	–
Macnair and Frank (2018b)	Islamic State	IS online media, transcripts of IS-produced videos	Sentiment analysis	<i>SentiStrength</i>	–
Mueller and Rauh (2018)	Prediction of armed conflict	Relevant newspaper articles	LDA topic modeling	<i>Python</i>	Alternative K_s^b
Arendt and Karadas (2017)	Mediated associations with Islam	German news coverage of Islam	Contrast Analysis of Semantic Similarity (CASS)	CASS software	Reference to Holtzman et al. (2011, 2015)
Medzhorsky et al. (2017)	Syrian Civil War	347 speeches from UN Security Council members of 37 states	Dictionary scaling with dynamic model	<i>R</i>	Qualitative checks

(continued)

Table 3 (continued)

Author(s)	Conflict	Data	Text mining technique	Software ^a	Validation ^a
Scharkow (2013)	Various news issues, including <i>crime</i> and <i>controversy</i>	933 hand-coded newspaper articles	Supervised learning	<i>Python</i>	Combined with hand-coding
Terman (2017)	Islamophobia	Media portrayals of Muslim women in US news media	STM topic modeling	<i>R</i>	–
Tingley (2017)	International conflicts	Survey data on bargaining between declining and rising power	STM topic modeling	<i>R</i>	Combined with experiment
Jacobi et al. (2016)	Nuclear technology	New York Times coverage of nuclear technology (1945–2013)	LDA topic modeling	<i>R</i>	Assessment of alternative K 's ^b , comparison with earlier deductive works
Törnberg and Törnberg (2016)	Islamophobia	Muslims in Swedish social media discourse (2000–2013)	LDA topic modeling	–	Combined with discourse analysis
O'Halloran et al. (2019)	Violent extremist discourse	IS propaganda magazine <i>Dabiq</i>	Dictionary analysis based on Wikipedia classifications, visualisation based on image processing algorithms	–	Qualitative analysis of text-image relations
Dowell et al. (2015)	Autocracies in crises	Texts of Fidel Castro, Zedong Mao, Hosni Mubarak	Coh-Matrix, LIWC dictionary	–	Reference to Tausczik et al. 2010
Lucas et al. (2015)	Islamic/Jihadi rhetoric	27 248 texts from Jihadists and non-Jihadists Muslim clerics	STM topic modeling	<i>R</i>	Network analysis and other visual tools
Mishler et al. (2015)	Ukrainian crisis	Ukrainian and Russian social media to study public opinion and detect events	STM topic modeling	<i>R</i>	Crosscheck with hand-coded sample

Burscher et al. (2014)	Conflicts described in news articles (among other frames)	Dutch front-page news articles (1995–2011)	Supervised learning	<i>Python</i>	Comparison of two models
Cohen et al. (2014)	Lone wolf terrorism	No empirical analysis	Discussion of different text analysis techniques to detect potential lone wolf terrorists in the Internet	–	–
Bonilla and Grimmer (2013)	Democracy and terrorism	51 766 US newspaper stories, major nightly news broadcasts (ABS, CBS, NBC)	LDA topic modeling	–	Survey data
Miller (2013)	Crime, unrest in Qing China (1722–1911)	Official records on violent crime and unrest ('the veritable records')	LDA topic modeling	–	–
Pennebaker (2011)	Language and terrorism	296 speeches, interviews, articles from four extremist groups	LIWC dictionary	–	Reference to Tausczik et al. 2010
Hart and Lind (2011)	Islamic activist rhetoric	327 Islamic texts compared with rhetoric of US politicians, protesters and preachers	Dictionary	DICTION software	–
Pennebaker (2011)	Islamist language styles	296 texts from four Arabic-speaking extremist groups with violent past	LIWC dictionary, function words	–	–
Stewart and Zhukov (2009)	Public debates in Russia over the use of force	7,920 public statements (1998–2008)	Expressed Agenda Model	–	Combined with supervised techniques

The primary aim of the table is to offer conflict researchers interested in text mining a first impression about topics under investigation, applied techniques, software used and validation tests

^aNOTE: Table elements with—signify that no specific information about used software or validation procedures was found in the respective article

^b *K* = Number of topics

References

- Albaugh, Q., Sevenans, J., Loewen, P. J., & Soroka, S. (2013). The automated coding of policy agendas: A dictionary-based approach. In *Proceedings of the 6th comparative agendas project (CAP)*, Antwerp.
- Arendt, F., & Karadas, N. (2017). Content analysis of mediated associations: An automated text-analytic approach. *Communication Methods and Measures*, 11(2), 1–16.
- Baden, C. & Tenenboim-Weinblatt, K. (2018). The search for common ground in conflict news research: Comparing the coverage of six current conflicts in domestic and international media over time. *Media, War and Conflict*, 11(1), 22–45.
- Bagozzi, B. E. & Berliner, D. (2018). The politics of scrutiny in human rights monitoring: Evidence from structural topic models of US state department human rights reports. *Political Science Research and Methods*, 6(4), 661–677. http://www.journals.cambridge.org/abstract_S2049847016000443.
- Bastos, M. & Mercea, D. (2018). Parametrizing Brexit: mapping Twitter political space to parliamentary constituencies. *Information, Communication & Society*, 21(7), 921–939.
- Ben-David, A. & Matamoros-Fernández, A. (2016). Hate speech and covert discrimination on social media: Monitoring the Facebook pages of extreme-right political parties in Spain. *International Journal of Communication*, 10, 1167–1193.
- Benoit, K., Nulty, P., Obeng, A., Wang, H., Lauderdale, B., & Lowe, W. (2018). *Quanteda Package*. <https://cran.r-project.org/web/packages/quanteda/quanteda.pdf>.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022. <http://ci.nii.ac.jp/naid/110009545970/>
- Boecking, B., Hall, M., & Schneider, J. (2015). Event prediction with learning algorithms—A study of events surrounding the Egyptian revolution of 2011 on the basis of micro blog data. *Policy and Internet*, 7(2), 159–184.
- Bonilla, T. & Grimmer, J. (2013, December). Elevated threat levels and decreased expectations: How democracy handles terrorist threats. *Poetics*, 41(6), 650–669.
- Bosc, T., Cabrio, E., & Villata, S. (2016). Tweeties squabbling : Positive and negative results in applying argument mining on social media. In *Computational Models of Argument* (pp. 21–32).
- Boumans, J. W. & Trilling, D. (2016). Taking stock of the toolkit. *Digital Journalism*, 4(1), 8–23.
- Bradley, M. M. & Lang, P. J. (1999). Affective norms for English words (ANEW): Instruction manual and affective ratings. *Technical Report C-1, the Center of Research in Psychophysiology*, 30(1), 25–36.
- Brintzenhoff, W. (2011). Automated language processing : Exploring the relationship of social media and conflict in a comparative analysis of Arabic social media and conflict events reported in news media. In *Proceedings of the International Studies Association International Conference, Montréal* (pp. 1–13).
- Burden, B. C. & Sanberg, J. N. (2003). Budget rhetoric in presidential campaigns from 1952 to 2000. *Political Behavior*, 25(2), 97–118.
- Burnap, P. & Williams, M. L. (2015). Cyber hate speech on twitter: An application of machine classification and statistical modeling for policy and decision making. *Policy and Internet*, 7(2), 223–242.
- Burscher, B., Odijk, D., Vliegthart, R., Rijke, M. de, & Vreese, C. H. de. (2014). Teaching the computer to code frames in news: Comparing two supervised machine learning approaches to frame analysis. *Communication Methods and Measures*, 8(3), 190–206.
- Burscher, B., Vliegthart, R., & De Vreese, C. H. (2015). Using supervised machine learning to code policy issues: Can classifiers generalize across contexts? *The ANNALS of the American Academy of Political and Social Science*, 659(1), 122–131. <http://ann.sagepub.com/content/659/1/122.abstract?rss=1>.
- Chang, J., Gerrish, S., Wang, C., & Blei, D. M. (2009). Reading tea leaves: How humans interpret topic models. *Advances in Neural Information Processing Systems* 22, 288–296. <http://www.umiacs.umd.edu/~jbg/docs/nips2009-rtl.pdf>

- Cohen, K., Johansson, F., Kaati, L., & Mork, J. C. (2014). Detecting linguistic markers for radical violence in social media. *Terrorism and Political Violence*, 26(1), 246–256.
- Denny, M. J. & Spirling, A. (2018). Text preprocessing for unsupervised learning: Why it matters, when it misleads, and what to do about it. *Political Analysis*, 26, 168–189. <https://www.ssrn.com/abstract=2849145>.
- DiMaggio, P. (2015). Adapting computational text analysis to social science (and vice versa). *Big Data & Society*, 2(2), 1–5.
- Dowell, N. M., Windsor, L. C., & Graesser, A. C. (2015). Computational linguistics analysis of leaders during crises in authoritarian regimes. *Dynamics of Asymmetric Conflict*, 9(01–03), 1–12. <http://www.tandfonline.com/doi/abs/10.1080/17467586.2015.1038286>
- Fawcett, P., Jensen, M. J., 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. *Journal of Public Policy*, 1–21.
- González-Bailón, S. & Paltoglou, G. (2015). Signals of public opinion in online communication: A comparison of methods and data sources. *The ANNALS of the American Academy of Political and Social Science*, 659(1), 95–107. <http://ann.sagepub.com/content/659/1/95.abstract?rss=1>.
- Greene, D., O’Callaghan, D., & Cunningham, P. (2014). How many topics? Stability analysis for topic models. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8724 LNAI(PART 1), 498–513.
- Greene, K. T., Park, B., & Colaresi, M. (2018). Machine learning human rights and wrongs: How the successes and failures of supervised learning algorithms can inform the Debate about information effects. *Political Analysis, Online First*. <https://doi.org/10.1017/pan.2018.11>
- Grimmer, J. & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, 21(3), 267–297.
- Hart, R. P. (2000). *Diction 5.0 User’s Manual*. London: Scolari Software, Sage Press.
- Hart, R. P. & Lind, C. J. (2011). The rhetoric of Islamic activism: A DICTION study. *Dynamics of Asymmetric Conflict: Pathways toward Terrorism and Genocide*, 4 (2), 113–125.
- Holtzman, N. S., Schott, J. P., Jones, M. N., Balota, D. A., & Yarkoni, T. (2011). Exploring media bias with semantic analysis tools: Validation of the contrast analysis of semantic similarity (CASS). *Behavior Research Methods*, 43(1), 193–200.
- Holtzman, N. S., Kwong, S., & Baird, K. L. (2015). Exploring political ideologies of senators with semantic analysis tools: Further validation of CASS. *Journal of Language and Social Psychology*, 34(2), 200–212.
- Jacobi, C., Atteveldt, W. van, & Welbers, K. (2016 Jan). Quantitative analysis of large amounts of journalistic texts using topic modelling. *Digital Journalism*, 4(1), 89–106. <http://www.tandfonline.com/doi/full/10.1080/21670811.2015.1093271>.
- Kananovich, V. (2018). Framing the Taxation-Democratization link: An automated content analysis of cross-national newspaper data. *International Journal of Press/Politics*, 23(2), 247–267.
- Kellstedt, P. M. (2000). Media framing and the dynamics of racial policy preferences. *American Journal of Political Science*, 44(2), 245. <http://www.jstor.org/stable/2669308?origin=crossref>.
- Laver, M. & Garry, J. (2000). Estimating policy positions from political texts. *American Journal of Political Science*, 44(3), 619–634.
- Lee, M. & Mimno, D. (2014). Low-dimensional embeddings for interpretable anchor-based topic inference. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1319–1328. <http://arxiv.org/abs/1711.06826>
- Loughran, T. & McDonald, B. (2010). When is a Liability not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *Journal of Finance*, 66(1), 35–65. <http://onlinelibrary.wiley.com/doi/10.1111/j.1540-6261.2010.01625.x/full>.
- Lucas, C., Nielsen, R. A., Roberts, M. E., Stewart, B. M., Storer, A., & Tingley, D. (2015). Computer-assisted text analysis for comparative politics. *Political Analysis*, 23(02), 254–277.
- Macnair, L. & Frank, R. (2018a). Changes and stabilities in the language of Islamic state magazines: A sentiment analysis. *Dynamics of Asymmetric Conflict: Pathways toward Terrorism and Genocide*, 11(2), 109–120. <http://doi.org/10.1080/17467586.2018.1470660>

- Macnair, L. & Frank, R. (2018b). The mediums and the messages: Exploring the language of Islamic State media through sentiment analysis. *Critical Studies on Terrorism*, 00(00), 1–20. <https://doi.org/10.1080/17539153.2018.1447226>
- Maerz, S. F. (2019). Simulating pluralism: The language of democracy in hegemonic authoritarianism. *Open Access in Political Research Exchange*. <https://doi.org/10.1080/2474736X.2019.1605834>.
- Maerz, S. F. & Schneider, C. Q. (2019). Comparing public communication in democracies and autocracies: automated text analyses of speeches by heads of government. *Quality and Quantity, Online First*. <http://link.springer.com/article/10.1007/s1135-019-00885-7>.
- Medzihorsky, J., Popovic, M., & Jenne, E. K. (2017). Rhetoric of civil conflict management: United Nations security council debates over the Syrian civil war. *Research and Politics*, 4(2), 1–10. <http://journals.sagepub.com/doi/10.1177/2053168017702982>.
- Miller, G. A. (1995). WordNet: A lexical database for english. *Communications of the ACM*, 38(11), 39–41.
- Miller, I. M. (2013). Rebellion, crime and violence in Qing China, 1722–1911: A topic modeling approach. *Poetics*, 41(6), 626–649. <http://dx.doi.org/10.1016/j.poetic.2013.06.005>.
- Mishler, A., Hefright, B., Paletz, S. B. F., Golonka, E., & Ford, A. (2015). Using structural topic modeling to study public opinion and detect events. *Conference Paper, International Conference on Human-Computer Interaction*. https://link.springer.com/chapter/10.1007/978-3-319-21380-4_108
- Montiel, C. J., Salvador, A. M. O., See, D. C., & De Leon, M. M. (2014). Nationalism in local media during international conflict: Text mining domestic news reports of the China-Philippines Maritime Dispute. *Journal of Language and Social Psychology*, 33(5), 445–464.
- Mueller, H. & Rauh, C. (2018). Reading between the lines: Prediction of political violence using newspaper text. *American Political Science Review*, 112(2), 358–375.
- Nelson, L. K. (2017). Computational grounded theory: A methodological framework. *Sociological Methods and Research*. <https://doi.org/10.1177%2F0049124117729703>.
- O'Halloran, K. L., Tan, S., Wignell, P., Bateman, J. A., Pham, D. S., Grossman, M., et al. (2019). Interpreting text and image relations in violent extremist discourse: A mixed methods approach for big data analytics. *Terrorism and Political Violence*, 31(3), 454–474.
- Payson Conflict Study Group. (2001). *A glossary on violent conflict terms and concepts used in conflict prevention, mitigation, and resolution in the context of disaster relief and sustainable development*. https://reliefweb.int/sites/reliefweb.int/files/resources/6C8E6652532FE542C12575DD0044F2D-USAID_may01.pdf
- Pennebaker, J. W. (2011). Using computer analyses to identify language style and aggressive intent: The secret life of function words. *Dynamics of Asymmetric Conflict: Pathways toward Terrorism and Genocide*, 4(2), 92–102.
- Pennebaker, J. W., Boyd, R. L., & Francis, M. E. (2001). *Linguistic Inquiry and Word Count (LIWC)*. Austin: Pennebaker Conglomerates. www.LIWC.net
- Proksch, S.-O., Lowe, W., Wäckerle, J., & Soroka, S. (2019). Multilingual sentiment analysis: A new approach to measuring conflict in legislative speeches. *Legislative Studies Quarterly*, 44(1), 97–131. <http://doi.wiley.com/10.1111/lsq.12218>
- Puschmann, C. (2018). *Inhaltsanalyse mit R*. <http://inhaltsanalyse-mit-r.de>
- Python. (2018). *Python software foundation. Python language reference*. <http://www.python.org/>
- R. (2018). *The R Core Team: A language and environment for statistical computing. R foundation for statistical computing*, Vienna, Austria. <http://www.r-project.org>
- Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., et al. (2014). Structural topic models for open-ended survey responses. *American Journal of Political Science*, 58(4), 1064–1082.
- Roberts, M. E., Stewart, B. M., & Airoldi, E. M. (2016). A model of text for experimentation in the social sciences. *Journal of the American Statistical Association*, 111(515), 988–1003. <http://dx.doi.org/10.1080/01621459.2016.1141684>
- Roberts, M. E., Stewart, B. M., & Tingley, D. (2018). STM: R package for structural topic models. *Journal of Statistical Software*. <http://arxiv.org/abs/1709.04553>

- Scharkow, M. (2013). Thematic content analysis using supervised machine learning: An empirical evaluation using German online news. *Quality and Quantity*, 47(2), 761–773.
- Schwemmer, C. (2018). Stminsights: A shiny application for inspecting structural topic models [Software-Handbuch]. <https://github.com/cschwem2er/stminsights> (R package version 0.3.0).
- Stewart, B. M. & Zhukov, Y. M. (2009). Use of force and civil–military relations in Russia: An automated content analysis. *Small Wars and Insurgencies*, 20(2), 319–343.
- Stone, P., Dunphy, D., Smith, M., & Ogilvie, D. (1966). *The general inquirer: A computer approach to content analysis*. Cambridge: MIT Press.
- Tausczik, Y. R. & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1), 24–54.
- Terman, R. (2017). Islamophobia and media portrayals of Muslim Women: A computational text analysis of US news coverage. *International Studies Quarterly*, 61(3), 489–502.
- Tingley, D. (2017). Rising power on the mind. *International Organization*, 71(S1), 165–188.
- Törnberg, A. & Törnberg, P. (2016). Muslims in social media discourse: Combining topic modeling and critical discourse analysis. *Discourse, Context and Media*, 13, 132–142. <http://dx.doi.org/10.1016/j.dcm.2016.04.003>
- Trilling, D. (2018). *Doing computational social science with Python: An introduction*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2737682
- van Atteveldt, W., Kleinnijenhuis, J., Ruigrok, N., & Schlobach, S. (2008). Good news or bad news? Conducting sentiment analysis on Dutch text to distinguish between positive and negative relations. *Journal of Information Technology and Politics*, 5(1), 73–94.
- Young, L. & Soroka, S. (2012). Affective news: The automated coding of sentiment in political texts. *Political Communication*, 29(2), 205–231.
- Zamith, R. & Lewis, S. C. (2015). Content analysis and the algorithmic coder: What computational social science means for traditional modes of media analysis. *The ANNALS of the American Academy of Political and Social Science*, 659(1), 307–318. <http://ann.sagepub.com/content/659/1/307.abstract?rss=1>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.