Automated Content Analysis

Valerie Hase

1 Introduction

Due to the rise in processing power, advancements in machine learning (Grimmer et al. 2021), and the availability of large text corpora online, the use of computational methods including automated content analysis (van Atteveldt und Peng 2018) has rapidly increased. Automated content analysis is applied and developed across disciplines such as computer science, linguistics, political science, economics and – increasingly – communication science (Hase et al. 2022). Recent pieces offer theoretical introductions to the method (Benoit 2020; Boumans and Trilling 2016; DiMaggio 2015; Grimmer and Stewart 2013; Günther and Quandt 2016; Manning and Schütze 1999; Quinn et al. 2010; Scharkow 2012; van Atteveldt et al. 2019; Wettstein 2016; Wilkerson and Casas 2017). Similarly, tutorials on how to conduct such analyses are readily available online (Puschmann 2019; Silge and Robinson 2022; Watanabe and Müller 2021; Welbers et al. 2017; Wiedemann and Niekler 2017).

Automated content analysis or "text as data" methods describe an approach in which the analysis of text is, to some extent, automatically conducted by machines. While automated analyses for other types of content, for example images (Webb Williams et al. 2020), have also been proposed more recently, this study will focus on text. In contrast to manual coding, text is not read and understood as one unit, but automatically broken

V. Hase (✉)
Department of Media and Communication, LMU Munich, München, Deutschland
E-Mail: valerie.hase@ifkw.lmu.de

F. Oehmer-Pedrazzi et al. (Hrsg.), *Standardisierte Inhaltsanalyse in der Kommunikationswissenschaft – Standardized Content Analysis in Communication Research,* https://doi.org/10.1007/978-3-658-36179-2_3

down to its "features", for example single words such as "she" or "say". The complexity of texts is then reduced further by converting text to numbers: Texts are often understood based on how often different features, for example unique words, occur. Computers use feature occurrences as manifest indicators to infer latent properties from texts (Benoit 2020), for example negativity or emotions. Importantly, manual coding is still part of most automated analyses: Humans may construct dictionaries to automatically look up features expressing sentiment, code sentiment in texts as training data on which algorithms are trained, or create a gold standard of manually annotated texts against which the results of automated analyses are compared (Song et al. 2020; van Atteveldt et al. 2019).

When using text as data approaches, readers should bear in mind important caveats and limitations. Human decisions lie at the core of "automated" content analyses and thus necessarily introduce certain degrees of freedom to these approaches. For example, researchers have to decide how to prepare text for analyses (Denny and Spirling 2018) or choose a method to infer latent concepts of interest (Nelson et al. 2021), which can heavily impact results. Also, text as data approaches are costly: Not only does it take considerable effort to decide on how to conduct which steps of the analysis and write code to execute them. Studies often rely on large sets of manually annotated texts for the training or validation of algorithms, which require time and money for manual coders. As automated content analyses aim to infer latent concepts, researchers should also note that the method necessarily includes uncertainty and error: It cannot grasp texts in their full complexity, similar to manual coding (Grimmer and Stewart 2013). As Grimmer and Stewart (2013, p. 269, capitalization by authors) put it: "All Quantitative Models of Language Are Wrong – But Some Are Useful".

Related to this, there is an ongoing debate about which variables can and should be measured automatically instead of relying on human coding (di Maggio 2015). It seems that the more complex the latent construct that should be inferred, the less suitable automated approaches. For example, formal features such as the use of hyperlinks in text (Günther and Scharkow 2014) or an article's publication date (Buhl et al. 2019) are easily detected automatically. Text as data approaches can also identify events that are being reported on across articles (Trilling and van Hoof 2020) and, as such, news chains (Nicholls and Bright 2019). However, recent studies have cast doubt on the performance of automated analyses for grasping more complex variables at the core of communication studies: When measuring evaluations or sentiment, human coding clearly outperforms machines (van Atteveldt et al. 2021). Similarly, studies on automated measurements of frames (Nichols and Culpepper 2021) or media bias (Spinde et al. 2021) do not warrant optimism that text as data approaches are applicable for any kind of text or even better than human coding. Thus, automated approaches do not replace human abilities to understand text. Rather, they amplify them (Grimmer and Stewart 2013; Nelson et al. 2021), as do computational methods in general (van Atteveldt and Peng 2018).

Emerging trends in the field include approaches that try to better model syntactic relationships in texts, e.g., evaluations concerning a specific actor (Fogel-Dror et al. 2019). Others aim to more accurately grasp the semantic meanings of features through word embeddings (Mikolov et al. 2013; Pennington et al. 2014, but for a discussion of potential biases see Bolukbasi et al. 2016). Studies also propose mixed methods approaches where computational methods and manual coding support each other, often in an iterative process (Lewis et al. 2013; Nelson 2020). Recently, semi-automated methods in which manual input is used as a starting point have emerged (Watanabe 2021). Studies have also introduced new ways of resourceful and cheap data collection such as crowdsourcing (Lind et al. 2017).

2 Common steps of analysis and research designs

Automated content analysis typically consists of the following four steps (Wilkerson and Casas 2017): (1) data collection, (2) data preprocessing, (3) data analysis, and (4) data validation.

(1) *Data collection*. First, large text corpora need to be obtained through structured databases such as Lexis Uni or other third party providers, Application Programming Interfaces (APIs) for data from social networks or newspapers, or by scraping websites (Possler et al. 2019; van Atteveldt et al. 2019). The collection of large amounts of textual data often involves legal problems due to copyright issues (Fuchsloch et al. 2019).

(2) *Data preprocessing*. In what is called preprocessing, texts are then prepared for automated analysis. Potential units of analysis might be whole articles/social media messages, but also single paragraphs or sentences. Preprocessing reduces text units to those features that are informative for detecting differences or similarities between different text units and dismisses features that are not. In every study, researchers have to decide which parts of text are informative and hence which of the following steps are important for their analysis. Not only are there no standard preprocessing steps (Benoit 2020) but the choice of preprocessing steps influences results (Denny and Spirling 2018; Scharkow 2012). Common steps include (1) the removal of boilerplate, for example URLS included in texts obtained via scraping. Next, (2) tokenization, where text is broken down to its features, is important. Oftentimes, these features are unigrams, i.e., single words, such as “he” or “and” in what is called a “bag-of-words” approach: The order or context of words is not taken into account. In “bag-of-word” approaches, the occurrence of a feature is what counts, independent of where in a given text the feature occurs or which features occur in close proximity to it (van Atteveldt et al. 2019). However, there are more informative ways of feature extraction than unigrams: Stoll, Ziegele and Quiring (2020), for example, include n-grams. These may describe bigrams, i.e., an order of two words, such as “he walks”, or trigrams, i.e., an order of three words, such as “and then he”. More meaningful n-grams are collocations, i.e., specific words that often co-occur and, in conjunction, have a different meaning. Statistically checking

for words that frequently co-occur or using Named Entity Recognition (NER), where names for persons, organizations or organizations are automatically detected, would for example lead to the unigrams "United" and "States" to be included as one feature, namely the collocation "United States". Some analyses also distinguish several meanings a feature may have through Part-of-Speech (PoS) tagging. For example, "novel" as a noun and "novel" as an adjective describe two very different things (Manning and Schütze 1999). Further preprocessing steps might include discarding punctuation (3) and capitalization (4). In addition, (5) features with little informative values are often deleted. Depending on the research question, these might include numbers, so-called "stop words" (often based on ready-made lists, including for example "and", "the"), or features occurring in almost every or almost no text in what is called relative pruning. Lastly, many analyses try to reduce complexity through (6) stemming or lemmatizing (the feature "analyzed", for example, becomes "analyz" with stemming and "analyze" with lemmatizing). In "bag-of-words" approaches, texts are finally (7) represented in a document-feature-matrix where rows identify the unit of analysis (e.g., an article, a paragraph, a sentence) and columns identify how often a feature occurs in this unit (e.g., how often the unigram "terrorist" occurs in the first, the second unit and so forth).

(3) *Data analysis*. While recent overviews have used various systematizations for different methods in the field of automated content analysis, many distinguish between (1) dictionary and rule-based approaches, (2) supervised machine learning, and (3) unsupervised machine learning. While (1) and (2) include deductive approaches where known categories are assigned to texts, (3) is more inductive as it explores unknown categories (Boumans and Trilling 2016; Grimmer and Stewart 2013; Günther and Quandt 2016).

Deductive Approaches: Assigning known categories to text

(a) Dictionary and rules-based approaches often simply count the occurrence of features. Studies for example analyze whether news coverage of Islam mentions the feature "terrorism" (Hoewe and Bowe 2021). More complex studies use feature lists, also called dictionaries, to look up uncivil expressions (Muddiman et al. 2019) or topics in texts (Guo et al. 2016). Two kinds of dictionaries need to be differentiated: "Off-the-shelf" dictionaries such as the General Inquirer (Stone et al. 1966) or the Linguistic Inquiry and Word Count LIWC (Tausczik and Pennebaker 2010) are ready-made dictionaries developed to be applied across text genres or topics. As Taboada (2016) cautions researchers, many "off-the-shelf" dictionaries were developed based on specific genres and topics, namely user reviews of consumer products. Research shows a lack of agreement between different "off-the-shelf" dictionaries and for their results to differ from manual coding (Boukes et al. 2020; van Atteveldt et al., 2021). For sentiment analysis, Boukes et al. (2020, p. 98) therefore stress that "scholars should be conscious of the weak performance of the off-the-shelf sentiment analysis tools". In contrast, "organic" dictionaries are inductively developed feature lists used to deductively assign known categories such as sentiment or topics to text units. As they are developed related

to the research question and the corpus at hand, they are tailored for a specific genre (e.g., social media texts or news articles), topic (e.g., texts concerning climate change or economic development), and concept of interest (e.g., negative sentiment or incivility). Although the construction of "organic" dictionaries is quite demanding, they oftentimes offer better results and should be preferred over "off-the-shelf" dictionaries (Boukes et al. 2020; Muddiman et al. 2019). However, both types of dictionaries still have general pitfalls in that they cannot easily handle negation, irony or polysemy, meaning that the same feature might have a completely different meaning depending on its context (Benoit 2020). They are also often tailored to English-language only (Lind et al. 2019).

(b) *Supervised machine learning* uses manually annotated training data from which classifiers learn how to categorize previously unknown data. The method is for example applied to classify texts concerning their topics (Scharkow 2012) or whether or not they contain incivility (Stoll et al. 2020). First, variables are coded by human coders to create a training data set. Next, classifiers use this training data to learn which independent variables (for example, the frequency of features such as "bad" and "catastrophe") predict the dependent variable (for example, negative sentiment). They then predict sentiment classifications for a previously unknown set of test data, i.e., texts researchers want to classify automatically (for a detailed overview of analysis steps see Barberá et al. 2021; Mirończuk and Protasiewicz 2018; Pilny et al. 2019). There is a plethora of classifiers that can be used, for example the Naive Bayes Classifier or Support Vector Machines (Scharkow 2012). Different classifiers can also be combined to ensembles. Supervised machine learning is not without limitations: Not only does the training data need to be of sufficient size, which can often mean that a considerable number of texts have to be coded manually. Researchers should also be cautious of strong dependencies of the classifier on the training data set, meaning the classifier works well for training data but poorly for test data. To avoid this, researchers often apply k-fold cross validation where the corpus is split into k groups. Then, each group is used as the test data once while the rest of the groups are used as training data without any overlaps between training and test data sets (Manning and Schütze 1999). Researchers should also test how generalizable their classifier is across contexts, meaning if it can accurately predict categories for new data with slightly different topics or text genres (Burscher et al. 2015).

Inductive Approaches: Exploring unknown categories in text

(c) *Unsupervised machine learning* takes a more inductive "bottom-up" approach as, in contrast to the previous approaches, categories are not previously known or fed to the model as training data. Instead, they are induced from the corpus (Boumans and Trilling 2016). If one is interested in categorizing texts concerning their main topics, for example, and has no assumptions as to which topics exist, unsupervised machine learning would be suitable.

The most prominent unsupervised machine learning approach is topic modeling (Blei et al. 2003). As a method to identify topics (Maier et al. 2018) and, as some argue, in combination with other methods even frames (Walter and Ophir 2019, but see Nicholls

and Culpepper 2021), the method has been of increasing interest. Topic modeling identifies the relative prevalence of topics in texts based on word co-occurrences. It assumes that documents can be represented as mixture of different latent topics that are themselves characterized by a distribution over words (Blei et al. 2003; Maier et al. 2018). In contrast to single-membership models such as k-means clustering (Grimmer and Stewart 2013), topic modeling therefore allows for multiple topics to occur in a text. Recent applications such as structural topic modeling also enable researchers to analyze how covariates – for example the year a text was published or its author – influence topic prevalence or its content (Roberts et al. 2014). While some settings such as the number of topics to be estimated need to be specified before running the model, topics themselves are generated without human supervision. While less resources have to be put towards running the model, testing the reliability and validity of results produced by unsupervised machine learning can be quite demanding. In the case of topic modeling, researchers should, for example, check how results vary when estimating different numbers of topics, whether topics are robust and reproducible across model runs, and whether they are coherent and meaningful (Maier et al. 2018; Roberts et al. 2016; Wilkerson and Casas 2017). In particular, choosing the number of topics the model should identify is a highly subjective process that will likely influence results.

(4) *Data validation*. One should not blindly trust the results of any automated method. Therefore, validation is a necessary step (Grimmer and Stewart 2013). For more deductive approaches such as dictionaries and supervised machine learning, validation is relatively straightforward: Researchers already know which categories of interest, for example negative sentiment, might be found. Hence, validity is reassured by comparing automated results, i.e., which texts were assigned which sentiment, to a benchmark. Oftentimes, this benchmark is manually annotated data as a gold standard, here describing which sentiment humans would assign. While this gold standard not necessarily implies the "true" value as human coding is quite erroneous (DiMaggio 2015) even if intercoder reliability is reassured, it indicates on what humans would *agree* for a text to be the "true" sentiment.

The most frequently reported indices for the validity of automated analyses are precision and recall (Song et al. 2020). *Precision* indicates how many articles predicted to contain negative sentiment according to the automated analysis actually contain negative sentiment according to the manual benchmark: How good is the model at not creating too many false positives? For example, a value of .8 implies that 80 % of all articles that do contain negative sentiment according to the automated classification actually contain negative sentiment according to the manual benchmark. However, 20 % were misclassified as containing negative sentiment and do, in fact, not. *Recall* indicates how many articles that actually contain negative sentiment were found: How good is our model at not creating too many false negatives? For example, a value of .8 implies that 80 % of all articles with negative sentiment were found by the automated approach. However, 20 % were not because they had been misclassified as not containing negative sentiment when they in fact did (Manning and Schütze 1999). However, many studies do not yet report

such validity tests (Song et al. 2020). Clear thresholds for what constitutes satisfactory values for these indices have not yet been agreed upon either – in contrast to intercoder reliability values for manual content analysis. Validity tests are also not very informative if results are unbalanced, meaning some categories – such as negative sentiment – have few true positives or true negatives. Given the uncertainty of quality thresholds, the question of "how good is good enough" (van Atteveldt 2008, p. 208) is still up for discussion.

The validation of unsupervised models is less direct. While studies argue that topic models, for example, can be validated by manually checking whether topics are coherent (Quinn et al. 2010) and can be differentiated from other topics (Chang et al. 2009, see Grimmer and Stewart 2013 for other approaches), there are no clear thresholds for what constitutes a valid model. Also, validity tests are reported even less often. Another issue are concerns about the reliability of these models. As Wilkerson and Casas (2017) summarize, unsupervised approaches are often instable, meaning that repeated estimations or different starting values lead to different results.

3 Analytical constructs employed in automated content analysis

Due to the interdisciplinarity of the method, automated content analysis has been used to measure a variety of constructs. For the field of communication science, studies often focus on four constructs of interest (see similarly Boczek and Hase 2020):

1. **Actors:** Many studies in the field of communication science use manual analysis to analyze how often actors, e.g., politicians or parties, are mentioned in texts (Vos and van Aelst 2018). Automated content analysis might be of massive help in this context. The recognition of so-called "named entities" (NER), including persons, organizations, or locations, has a long tradition in computer science. While different approaches have been discussed and the correct recognition of named entities is not yet solved (Marrero et al. 2013), studies have introduced potential approaches to our field. Recent analyses for example use rule-based approaches and dictionaries (Lind and Meltzer 2021; van Atteveldt 2008), machine learning (Burggraaff and Trilling 2020), or combinations of these methods (Fogel-Dror et al. 2019) to automatically classify named entities and often entity-related sentiment in text. This already indicates why these approaches might be of interest: Not only can we automatically count names of entities mentioned in text. We can also measure how different entities relate to each other, e.g., who talks about whom (van Atteveldt 2008), and sentiment concerning specific actors, e.g., how an entity is evaluated (Fogel-Dror et al. 2019).

2. **Sentiment or Tone:** Many studies are interested less in entity-related sentiment and more in the general sentiment or tone of news, for example for economic (Boukes et al. 2020) or political coverage (Young and Soroka 2012). A plethora of overview articles deliver introductions to such approaches which are often discussed in the context of sentiment analysis (Stine 2019; Taboada 2016). Sentiment analysis has developed from relying on dictionaries to using machine learning to applying deep learning and neural networks. Stine (2019) shows that the method delivered better performances with each turn in methods: While off-the-shelf dictionaries deliver insufficient results (Boukes et al. 2020) and organic dictionaries tailored to the genre, topic and concept of interest in one's study are recommended instead (Muddiman et al. 2019), supervised approaches seem to offer better results than at least off-the-shelf dictionaries (Barberá et al. 2021; González-Bailón and Paltoglou 2015). However, artificial neural networks can also be a suitable approach, especially for unbalanced data (Moraes et al. 2013; Stine 2019) and have already been applied in communication science (Rudkowsky et al. 2018). In sum, machine learning approaches in general might be better suited to analyze sentiment than dictionaries (Barberá et al. 2021). However, almost all of these methods still fall short of human coding (van Atteveldt et al. 2021).
3. **Topics:** Many analyses are interested in topics, i.e., what is being talked about in texts. A plethora of methods has been applied to analyze topics: Many studies use supervised machine learning in the form of topic modeling (Blei et al. 2003; Maier et al. 2018; Quinn et al. 2010) while others have applied supervised machine learning (Burscher et al. 2015; Scharkow 2012) or dictionaries (Guo et al. 2016). Related to these studies, Trilling and van Hoof (2021) have proposed and compared different methods to detect events in text**.** While dictionaries seem to perform slightly worse than unsupervised machine learning (Guo et al. 2016), choosing a suitable method depends more on whether researchers already know which topics may appear (Grimmer and Stewart 2013). Supervised learning or dictionaries are more appropiate if a study is interested in identifying a set of predetermined topics. If these are unknown, (structural) topic modeling may be a better fit (Roberts et al. 2014).
4. **Frames:** Lastly, many communication scholars are interested not only in what is being talked about in texts but also how issues are being talked about, in particular framing as the selection and salience of specific aspects (Entman 1993). Recent studies have tried to detect frames based on computational methods, mostly by analyzing topics using unsupervised machine learning. They then map similar topics to overarching frames using network analysis and community detection algorithms (Walter and Ophir 2019) or cluster analysis (van der Meer et al. 2019) in a second step. Others have applied supervised machine learning (Burscher et al. 2014) or compared a range of methods (Nicholls and Culpepper 2021). However, researchers should refrain from presuming that constructs identified through computational methods can (always) be called frames, especially based on unsupervised approaches (Nicholls and Culpepper 2021; van Atteveldt et al. 2014).

4 Research desiderata

Automated content analysis has gained in importance across disciplines, including communication science. In pace with rising computational power, it has transformed the ways in which we think about and approach analyses of text. However, standards for how to conduct these analyses are still evolving. Moreover, which method and analyses steps are most suitable for a specific study depends on the data and research question at hand (Grimmer and Stewart 2013). In reality, the availability of computational power, manually annotated data or a researchers' coding and statistical knowledge often influence such choices. Many departments in the field of communication science do not yet offer courses on statistics or programming that are necessary for communication scientists to fully understand and apply these methods (Boczek and Hase 2020).

Furthermore, the lack of available methods outside of bag-of-word approaches represents a research desideratum. Especially when dealing with more complex questions above and beyond how often a certain word or actor is mentioned in a given text, for example relationships between actors, studies need to more strongly consider syntactic relationships. Approaches for this have already been proposed (Fogel-Dror et al. 2019; van Atteveldt 2008), but most analyses still rely on the quite unrealistic "bag-of-words" assumption.

Another ongoing issue are concerns about the reliability and validity of computational methods (Nelson 2019), which are often neither tested nor reported. Uncertainty and error are an inherent part of automated analyses, similar to manual content where intercoder reliability reflects disagreement between individual coders. Given that studies using manual content analysis almost always need to report intercoder values for publication, similar thresholds for what constitutes a reliable and valid automated content analysis should be developed and be made mandatory for publication of automated analyses. Also, when deciding between manual and automated approaches, innovativeness should not outweigh the reliability and validity of results. While computational methods are often seen as a (methodological) advancement, they still have to satisfy essential validity and reliability thresholds for scholars to trust their results. In conclusion: Researchers should not choose computational methods over existing approaches simply because they seem more innovative.

The biggest question, however, is as follows: Even if we measure latent constructs such as topics, frames, or sentiment through automated content analysis – do we actually capture things that are relevant for theories and frameworks within communication science? Take topic modeling: There is an ongoing discussion about what topics mean (Maier et al. 2018). Are topics simply issues discussed in the news (van Atteveldt et al. 2014) or, if clustered, may they be interpreted as frames (Walter and Ophir 2019)? In other words, what do we gain by measuring topics in the news? Among other things, the shift to computational social science brings forward rigorous demands not only for statistical analysis or research designs but theory building (Peng et al. 2019). And while computational methods may inspire such (Waldherr et al. 2021), the status quo

leaves further fruit of thought not only for methodological advances, but also for how computational methods might change existing and push new theories in communication science.

Relevant Variables in DOCA – Database of Variables for Content Analysis

Frames: https://doi.org/10.34778/1c
Actors: https://doi.org/10.34778/1b
Sentiment/tone: https://doi.org/10.34778/1d
Topics: https://doi.org/10.34778/1e

References

Barberá, P., Bodystun, A. E., Linn, S., McMahon, R., & Nagler, J. (2021). Automated text classification of news articles: A practical guide. *Political Analysis*, *29*(1), 19–42.

Benoit, K. (2020). Text as data: An overview. In L. Curini & R. Franzese (Eds.), *The SAGE handbook of research methods in political science and international relations* (pp. 461–497). London: Sage.

Blei D.M., Ng A.Y., & Jordan, M.I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research,* 3, 993–1022.

Boczek, K., & Hase, V. (2020). Technische Innovation, theoretische Sackgasse? Chancen und Grenzen der automatisierten Inhaltsanalyse in Lehre und Forschung. In J. Schützeneder, K. Meier, & N. Springer (Eds.), Neujustierung der Journalistik/Journalismusforschung in der digitalen Gesellschaft: Proceedings zur Jahrestagung der Fachgruppe Journalistik/Journalismusforschung der Deutschen Gesellschaft für Publizistik- und Kommunikationswissenschaft 2019, Eichstätt (pp. 117–128). doi:https://doi.org/10.21241/ssoar.70828.

Bolukbasi, T., Chang, K.-W., Zou, J., Saligrama, V., & Kalai, A (2016). Quantifying and reducing stereotypes in word embeddings. Retrieved from: https://arxiv.org/abs/1606.06121.

Boukes, M., van de Velde, B., Araujo, T., & Vliegenthart, R. (2020). What's the tone? Easy doesn't do it: Analyzing performance and agreement between off-the-shelf sentiment analysis tools. *Communication Methods and Measures*, *14*(2), 83–104.

Boumans, J. W., & Trilling, D. (2016). Taking stock of the toolkit: An overview of relevant automated content analysis approaches and techniques for digital journalism scholars. *Digital Journalism*, *4*(1), 8–23.

Buhl, F., Günther, E., & Quandt, T. (2019). Bad news travels fastest: A computational approach to predictors of immediacy in digital journalism ecosystems. *Digital Journalism*, *7*(7), 910–931.

Burggraaff, C., & Trilling, D. (2020). Through a different gate: An automated content analysis of how online news and print news differ. *Journalism*, *21*(1), 112–129.

Burscher, B., Odijk, D., Vliegenthart, R., de Rijke, M., de Vreese, C.H. (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., Vliegenthart, 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.

Chang, J., Boyd-Graber, J., Wang, C., Gerrish, S., & Blei, D. M. (2009). Reading tea leaves: how humans interpret topic models. Paper presented at the *Neural Information Processing Systems*

2009. Retrieved from https://papers.nips.cc/paper/3700-reading-tea-leaves-how-humans-interpret-topic-models.pdf.

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*(2), 168–189.

DiMaggio, P. (2015). Adapting computational text analysis to social science (and vice versa). *Big Data & Society*, *2*(2), 1–5.

Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication, 43*(4), 51–58.

Fogel-Dror, Y., Shenhav, S. R., Sheafer, T., & van Atteveldt, W. (2019). Role-based association of verbs, actions, and sentiments with entities in political discourse. *Communication Methods and Measures*, *13*(2), 69–82.

Fuchsloch, S., von Nordheim, G., & Boczek, K. (2019). Unlocking digitized public spheres: Research opportunities and legal challenges in the use of text mining for content analysis. In C. Peter, T. K. Naab, & R. Kühne (Eds.), *Measuring media use and exposure: Recent developments and challenges* (Vol. 14, pp. 266–296). Cologne: Herbert von Halem Verlag.

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.

Grimmer, J., Roberts, M. E., & Stewart, B. M. (2021). Machine learning for social science: An agnostic approach. *Annual Review of Political Science*, *24*(1), 395–491.

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.

Günther, E., & Scharkow, M. (2014). Recycled media. An automated evaluation of news outlets in the twenty-first century. *Digital Journalism*, *2*(4), 524–541.

Günther, E., & Quandt, T. (2016). Word counts and topic models: Automated text analysis methods for digital journalism research. *Digital Journalism*, *4*(1), 75–88.

Guo, L., Vargo, C. J., Pan, Z., Ding, W., & Ishwar, P. (2016). Big social data analytics in journalism and mass communication: Comparing dictionary-based text analysis and unsupervised topic modeling. *Journalism & Mass Communication Quarterly*, *93*(2), 332–359.

Hase, V., Mahl, D., & Schäfer, M.S. (2022). Der „Computational Turn": ein „interdisziplinärer Turn"? Ein systematischer Überblick zur Nutzung der automatisierten Inhaltsanalyse in der Journalismusforschung. *Medien & Kommunikationswissenschaft, 70*(1–2), 60–78.

Hoewe, J., & Bowe, B. J. (2021). Magic words or talking point? The framing of 'radical Islam' in news coverage and its effects. *Journalism*, *22*(4), 1012–1030.

Lewis, S. C., Zamith, Rodrigo, & Hermida, A. (2013). Content analysis in an era of big data: A hybrid approach to computational and manual methods. *Journal of Broadcasting & Electronic Media, 57*(1), 34–52.

Lind, F., Eberl, J.-M., Heidenreich, T., & Boomgarden, H.G. (2019). When the journey is as important as the goal: A roadmap to multilingual dictionary construction. *The International Journal of Communication, 13,* 4000–4020.

Lind, F., Gruber, M., & Boomgaarden, H. G. (2017). Content analysis by the crowd: Assessing the usability of crowdsourcing for coding latent constructs. *Communication Methods and Measures*, *11*(3), 191–209.

Lind, F., & Meltzer, C. E. (2021). Now you see me, now you don't: Applying automated content analysis to track migrant women's salience in German news. *Feminist Media Studies, 21*(6), 923–940.

Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., Pfetsch, B., Heyer, G., Reber, U., Häussler, T., Schmid-Petri, H., & Adam, S. (2018). Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. *Communication Methods and Measures*, *12*(2–3), 93–118.

Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing*. Cambridge: MIT Press.

Marrero, M., Urbano, J., Sánchez-Cuadrado, S., Morato, J., & Gómez-Berbís, J. M. (2013). Named entity recognition: Fallacies, challenges and opportunities. *Computer Standards & Interfaces*, *35*(5), 482–489.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. Retrieved from http://arxiv.org/abs/1301.3781.

Mirończuk, M. M., & Protasiewicz, J. (2018). A recent overview of the state-of-the-art elements of text classification. *Expert Systems with Applications*, *106*, 36–54.

Moraes, R., Valiati, J. F., & Gavião Neto, W. P. (2013). Document-level sentiment classification: An empirical comparison between SVM and ANN. *Expert Systems with Applications*, *40*(2), 621–633.

Muddiman, A., McGregor, S. C., & Stroud, N. J. (2019). (Re)claiming our expertise: Parsing large text corpora with manually validated and organic dictionaries. *Political Communication*, *36*(2), 214–226.

Nelson, L. K. (2020). Computational grounded theory: A methodological framework. *Sociological Methods & Research*, *49*(1), 3–42.

Nelson, L. K. (2019). To measure meaning in big data, don't give me a map, give me transparency and reproducibility. *Sociological Methodology*, *49*(1), 139–143.

Nelson, L. K., Burk, D., Knudsen, M., & McCall, L. (2021). The future of coding: A comparison of hand-coding and three types of computer-assisted text analysis methods. *Sociological Methods & Research*, *50*(1), 202–237.

Nicholls, T., & Bright, J. (2019). Understanding news story chains using information retrieval and network clustering techniques. *Communication Methods and Measures*, *13*(1), 43–59.

Nicholls, T., & Culpepper, P.D. (2021). Computational identification of media frames: Strengths, weaknesses, and opportunities. *Political Communication*, *38*(1–2), 159–181.

Niekler, A., & Wiedemann, G. (2019). *Text mining for humanists and social scientists in R*. Retrieved from https://tm4ss.github.io/docs/index.html.

Peng, T.-Q., Liang, H., & Zhu, J. J. H. (2019). Introducing computational social science for Asia-Pacific communication research. *Asian Journal of Communication*, *29*(3), 205–216.

Pennington, J., Socher, R., & Manning, C. (2014). GloVe: Global vectors for word representation. Retrieved via https://nlp.stanford.edu/projects/glove/.

Pilny, A., McAninch, K., Slone, A., & Moore, K. (2019). Using supervised machine learning in automated content analysis: An example using relational uncertainty. *Communication Methods and Measures*, *13*(4), 287–304.

Possler, D., Bruns, S., & Niemann-Lenz, J. (2019). Data is the new oil – but how do we drill it? Pathways to access and acquire large data sets in communication science. *The International Journal of Communication, 13*, 3894–3911.

Puschmann, C. (2019). *Automatisierte Inhaltsanalyse mit R*. Retrieved from http://inhaltsanalyse-mit-r.de.

Quinn, K. M., Monroe, B. L., Colaresi, M., Crespin, M. H., & Radev, D. R. (2010). How to analyze political attention with minimal assumptions and costs. *American Journal of Political Science*, *54*(1), 209–228.

Roberts, M. E., Stewart, B. M., & Tingley, D. (2016). Navigating the local modes of big data: The case of topic models. In R. M. Alvarez (Ed.), *Computational Social Science* (pp. 51–97). Cambridge: Cambridge University Press.

Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., Albertson, B., & Rand, D. G. (2014). Structural topic models for open-ended survey responses. *American Journal of Political Science*, *58*(4), 1064–1082.

Rudkowsky, E., Haselmayer, M., Wastian, M., Jenny, M., Emrich, Š., & Sedlmair, M. (2018). More than bags of words: Sentiment analysis with word embeddings. *Communication Methods and Measures*, *12*(2–3), 140–157.

Scharkow, M. (2012). *Automatische Inhaltsanalyse und maschinelles Lernen*. Berlin: epubli.

Silge, J., & Robinson, D. (2022). *Text mining with R*. Retrieved from https://www.tidytextmining.com.

Song, H., Tolochko, P., Eberl, J.-M., Eisele, O., Greussing, E., Heidenreich, T., Lind, F., Galyga, S., & Boomgarden, H. G. (2020). In validations we trust? The impact of imperfect human annotations as a gold standard on the quality of validation of automated content analysis. *Political Communication*, *37*(4), 550–572.

Spinde, T., Rudnitckaia, L., Mitrović, J., Hamborg, F., Granitzer, M., Gipp, B., Donnay, K. (2021): Automated identification of bias inducing words in news articles using linguistic and context-oriented features. *Information Processing & Management*, *58*(3), 102505.

Stine, R. A. (2019). Sentiment analysis. *Annual Review of Statistics and Its Application*, *6*, 287–308.

Stoll, A., Ziegele, M., & Quiring, O. (2020). Detecting impoliteness and incivility in online discussions. Classification approaches for german user comments. *Computational Communication Research*, *2*(1), 109–134.

Stone, P. J., Dunphy, D. J., & Smith, M. S. (1966). *The general inquirer: A computer approach to content analysis*. Cambridge: M.I.T. Press.

Taboada, M. (2016). Sentiment analysis: An overview from linguistics. *Annual Review of Linguistics*, *2*, 325–347.

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.

Trilling, D., & van Hoof, M. (2020). Between article and topic: News events as level of analysis and their computational identification. *Digital Journalism, 8*(10), 1317–1337.

van Atteveldt, W. (2008). *Semantic network analysis: Techniques for extracting, representing and querying media content*. Charleston: BookSurge.

van Atteveldt, W., & Peng, T.-Q. (2018). When communication meets computation: Opportunities, challenges, and pitfalls in computational communication science. *Communication Methods and Measures*, *12*(2–3), 81–92.

van Atteveldt, W., van der Velden, M. A. C. G., & Boukes, M. (2021). The validity of sentiment analysis: Comparing manual annotation, crowd-coding, dictionary approaches, and machine learning algorithms. *Communication Methods and Measures*, *15*(2), 121–140.

van Atteveldt, W., Welbers, K., Jacobi, C., & Vliegenthart, R. (2014). *LDA models topics... But what are 'topics'?* Retrieved from http://vanatteveldt.com/wp-content/uploads/2014_vanatteveldt_glasgowbigdata_topics.pdf.

van Atteveldt, W., Welbers, K., & van der Velden, M. (2019). Studying political decision making with automatic text analysis. In W. R. Thompson (Ed.), *Oxford Research Encyclopedia of Politics*. Oxford University Press.

van der Meer, T. G. L. A., Kroon, A. C., Verhoeven, P., & Jonkman, J. (2019). Mediatization and the disproportionate attention to negative news: The case of airplane crashes. *Journalism Studies*, *20*(6), 783–803.

Vos, D., & van Aelst, P. (2018). Does the political system determine media visibility of politicians? A comparative analysis of political functions in the news in sixteen countries. *Political Communication*, *35*(3), 371–392.
Waldherr, A., Geise, S., Mahrt, M., Katzenbach, C., & Nuernbergk, C. (2021). Toward a stronger theoretical grounding of computational communication science: How macro frameworks shape our research agendas. *Computational Communication Research*, *3*(2), 1–28.
Walter, D., & Ophir, Y. (2019). News frame analysis: an inductive mixed-method computational approach. *Communication Methods and Measures*, *13*(4), 248–266.
Watanabe, K. (2021). Latent semantic scaling: A semisupervised text analysis technique for new domains and languages. *Communication Methods and Measures*, *15*(2), 81–102.
Watanabe, K., & Müller, S. (2021). Quanteda tutorials. Retrieved from https://tutorials.quanteda.io.
Webb Williams, N., Casas, A., & Wilkerson, J. D. (2020). Images as data for social science research. Cambridge: Cambridge University Press.
Welbers, K., van Atteveldt, W., & Benoit, K. (2017). Text analysis in R. *Communication Methods and Measures*, *11*(4), 245–265.
Wettstein, M. (2016). *Verfahren zur computerunterstützten Inhaltsanalyse in der Kommunikationswissenschaft*. Retrieved from http://opac.nebis.ch/ediss/20162838.pdf.
Wilkerson, J., & Casas, A. (2017). Large-scale computerized text analysis in political science: Opportunities and challenges. *Annual Review of Political Science*, *20*, 529–544.
Young, L., & Soroka, S. (2012). Affective news: The automated coding of sentiment in political texts. *Political Communication*, *29*(2), 205–231.

Valerie Hase is a Research Assistant at the Department of Media and Communication, LMU Munich. Her research focuses on crisis and conflict communication, terrorism and text as data/computational social science.