Topic Modeling

Melanie Althage

https://orcid.org/0000-0001-5233-1061

Abstract Topic Modeling is a method used in the Digital Humanities to examine the thematic structure of large collections of texts. This chapter offers an introduction to its methodological foundations. In addition to an overview of various Topic Modeling algorithms and their respective fields of application, the article focuses on central workflow aspects, such as the preparation of the text data (preprocessing) and evaluation of the modeling results. The aim is to provide a solid basis for the critically reflected use of Topic Modeling in theological research.*

Keywords Topic Modeling, Text Mining, Quantitative Text Analysis, Machine Learning, Natural Language Processing, Blended Reading, Distant Reading

1. Introduction

In a 2006 article, Gregory Crane asks: "What do you do with a million books" (Crane 2006)? This question has become increasingly relevant with the readily available number of digital sources (see also Stulpe & Lemke 2016, 18). However, a significant portion of these sources is only weakly structured, making it difficult to retrieve the information they contain. In what ways can this wealth of information and potential knowledge be effectively explored and made useful for research purposes? One answer to this question is provided by Topic Modeling. Topic Modeling is a clustering algorithm that uses linguistic patterns to structure large text corpora thematically and make them searchable. If one assumes that themes or content-related concepts are expressed through a specific set of terms that frequently co-occur in different historical sources, such automated pattern recognition methods can provide valuable contributions to research.

In the Digital Humanities and historical sciences, Topic Modeling has established itself as a versatile tool for a wide range of research questions. The method enables the analysis of research trends in academic journals (Mimno 2012; Wehrheim 2019; Wehrheim et al. 2022), the investigation of discourse structures in various publications

* This chapter, including quotations in foreign languages, was translated from German by Brandon Watson.

Melanie Althage: Topic Modeling, in: Christopher A. Nunn and Frederike van Oorschot (eds.): Compendium of Computational Theology 1. Introducing Digital Humanities to Theology, Heidelberg: heiBOOKS, 2024, S. 245–266. DOI: https://doi.org/10.11588/heibooks.1521.c21952

(Völkl et al. 2022; Bunout & von Lange 2019) or the positioning of Digital Humanities as a discipline in comparison to other fields (Luhmann & Burghardt 2021). In the context of theology, the method is also increasingly being used. Christopher A. Nunn, for example, presented Topic Modeling in his study as part of a broader distant reading approach and used the *DARIAH-DE Topics Explorer* (Simmler et al. 2019), a user-friendly software, to shed light on ethical topics in the letters of Augustine of Hippo (Nunn 2022). Mark Graves examined the model-theoretical and mathematical-computational aspects of Topic Modeling for his study on the moral theology of Thomas Aquinas. He demonstrated how the method can be used to analyze complex moral and theological concepts in their various facets and subsequently investigate their influence on papal encyclicals (Graves 2022).

To encourage further studies in theology, this article aims to provide a critical and reflective introduction to the method and workflow of Topic Modeling as well as its many variants and configuration possibilities. It outlines not only the potentials, but also the limitations and challenges that need to be considered when using this method in the research process. The article first outlines the basic concept of Topic Modeling. Then, the article provides an overview of various algorithms and their usage. A detailed description of the mathematical principles behind the individual methods is deliberately omitted; for in-depth information, please consult the relevant specialist literature. Finally, the article discusses the central aspects of data preparation and evaluation of the modeling results. The aim is to provide a foundation and initial orientation for the application of Topic Modeling in theological research.[1]

1 The exemplary topics presented in Figures 1–3 are derived from the German-language book reviews published on the specialist communication portal H-Soz-Kult (https://www.hsozkult.de/, accessed: 19 July 2024) between 1996 and June 2019 (15,103 reviews with approximately 18 million words). The selected topics are taken from a model comprising a total of 80 topics, created using the *Latent Dirichlet Allocation* (LDA; Blei et al. 2003) algorithm implemented in the *MALLET* software (McCallum 2002) via the Python wrapper in *Gensim* (Řehůřek & Sojka 2010) as part of the author's ongoing dissertation project, which is provisionally titled: “Mining the Historian's Web – A Method-critical Reflection on Quantitative Methods for the Analysis of Born-Digital Sources Using the Example of Historical Specialist Communication”. They are based on an earlier phase of the project. The topics in Table 1, which are used as examples for illustration, are in turn based on selected German-language funeral sermons from the 17th century (299 with approximately 3 million words). These sources were digitized as part of the *German Research Foundation* (DFG)-funded project *AEDit Frühe Neuzeit* in cooperation with the *German Text Archive* (*Deutsches Textarchiv,* DTA). They were prepared for machine-readability in accordance with the DTA transcription guidelines. For the “AEDit Frühe Neuzeit” sub-corpus see: Deutsches Textarchiv. Grundlage für ein Referenzkorpus der neuhochdeutschen Sprache. Edited by the Berlin-Brandenburgische Akademie der Wissenschaften, Berlin 2024, URL: www.deutschestextarchiv.de/search/metadata?corpus=aedit. Accessed: 19 July 2024. For comparability, these models were also generated using the MALLET wrapper from *Gensim.*

2. Methodological Basis

Topic Modeling is a method of Text Mining that aims to access and understand the content of extensive text corpora (for an introductory overview: Blei 2012 a; b; Brett 2012). Unlike classification algorithms where categories are explicitly specified *(supervised machine learning),* Topic Modeling is based on a generative probabilistic modeling process *(unsupervised machine learning).* In this process, the *categories* or *topics* are derived directly from the data. The method resembles traditional indexing practices that have been used since the 18th century to efficiently access certain text units; however, it differs in its approach: instead of fixed keywords, heterogeneous word clusters are generated through probabilistic calculations (Piper 2018, 66–75; see Fig. 1 for an example).

According to *Latent Dirichlet Allocation* (LDA; Blei et al. 2003), the classic Topic Modeling process assumes that the documents of an extensive corpus are made up of different proportions of a fixed set of themes. Furthermore, these themes can be reconstructed as latent, i.e., hidden, linguistic structures or patterns from the text data via the generation of topics (see Blei 2012b, 78–82 for details on the assumptions and the modeling process). To illustrate this, let us assume that we have access to a digitally available library of theological works on Christianity, the content categorization of which via keywords has been lost. The works could contain references to the Trinity, salvation, ethics and morality, as well as biblical exegesis. Topic Modeling enables a reconstruction of these latent content categories. However, Topic Modeling does not generate specific keywords, such as "Trinity," but groups of words (e.g., "God, Jesus, Spirit, Father, Son, Holy, Trinity …") that occur together statistically often in the individual documents. The aim is therefore to identify groups of words which, by interpreting their composition, provide an overview of the content structure of the library and its individual works.

Fig. 1 is an exemplary selection of topics in the history of religion for book reviews published on HSoz-Kult; visualization form: word clouds with a weighting of the words according to the relevance for the topic.

In the first step, each word in the works is randomly assigned to a topic. Likewise, a random combination of topics is assigned to each work. In the next step, these initial assignments are checked. The frequency and co-occurrence of a word with other words are used to evaluate whether the current topic assignment is appropriate or whether the word aligns more with a different word cluster. The same method applies to the individual documents: a work in which the words “Jesus”, “holy”, “grace”, “forgiveness”, “sin” and “redemption” occur frequently could, for example, be about the concept of salvation, but was perhaps initially assigned to the topic of “Trinity”. These false assignments are then updated.[2] This process is repeated many, often thousands of times (so-called iterations), until the corpus is structured in a “meaningful” way. In this case, meaningful means that hardly any assignment changes are necessary because the model has stabilized.[3]

Finally, the method produces a statistical model of the library that enables the assignment of individual works to theological topics and ensures efficient orientation within the corpus. This model is represented by two forms of *outputs.* On the one hand, the sources are usually represented as a *document-topic matrix*, i.e., a table documenting the topic weightings for each document. On the other hand, a *topic-word matrix* is generated analogously, breaking down the percentage weighting of the individual words for the individual topics (see also Althage 2022, 260f.). In so doing, one can abstract from the specific works and extract certain relevant features of the individual texts (i.e., the statistically relevant patterns in language use) as numerical representations, with the aim of understanding the content of the corpus.

Quantitative text analysis methods such as Topic Modeling, which process texts in the form of numerical representations, may initially seem unfamiliar to fields typically engaged in qualitative text-hermeneutic research, such as theology. However, with their macro-analytical approach (cf. Jockers 2013; Graham et al. 2016) these quantitative forms of analyses offer new perspectives on objects of research. There are numerous conceivable applications for theology. Above all, this method is fruitful for analyzing dominant themes, discourses, or concepts – for instance, in sermons, letters, works of the Church Fathers, or scholarly literature. These methods allow one to examine how focal points change over time as well as how different themes or concepts relate to one another. By analyzing texts from different religious groups or authors, one could identify differences and similarities in theological perspectives; various types of texts could also be examined in terms of their linguistic and thematic characteristics.

2 This “assignment” of a topic to a document is expressed as a probability value, which says something about the likelihood of this word cluster occurring in the document or overall corpus.

3 With most Topic Modeling algorithms, one must determine how many topics to be generated for a corpus and in how many repetitions (iterations); one is advised to try different configurations depending on the corpus and the expected diversity of topics (see also section 4).

These possible applications (see also Althage 2022, 259f.) arise from Topic Modeling's ability to process texts as data and thus perform a systematic and scalable analysis. In traditional research contexts, samples or case studies are typically used for an exemplary investigation. Conversely, computational methods can be applied to arbitrarily large source corpora with sufficient computing capacity, thereby also extending the periods of investigation. Given the human cognitive process, extensive corpora are difficult to analyze with consistent examination and relevance criteria, thus what is extracted from the sources develops dynamically and is influenced by a variety of factors (keyword: hermeneutic circle). On the other hand, the computer easily processes very large amounts of data systematically and consistently. The generated topics are provided solely from the data and are not based on categories previously defined based on certain presuppositions.[4] An additional advantage is that Topic Modeling can be applied to any language and therefore to any source data. Through a systematic approach, Topic Modeling enables an in-depth analysis of not just one, but thousands of documents, allowing for the identification and interpretation of hidden thematic structures, thereby gaining a deeper understanding of the particularities of the research object and challenging preconceived assumptions.

3. Topics: Definition and Epistemological Limits

In view of the previously outlined scope of application of Topic Modeling, the term *topic* needs to be defined more precisely. A clearly defined term will help to prevent any misconceptions about the knowledge potential of this method. As with many Text Mining methods, Topic Modeling is mainly based on counting word frequencies. In this context, a topic is a probability distribution across the vocabulary of the text collection that describes the co-occurrence of certain words (Blei 2012b, 78). Although the term "topic" may suggest a resemblance to "Topik" or "Topoi" (Piper 2018, 66–75; Horstmann 2018, 4–7), the term does not carry any epistemological implications beyond the probabilities of co-occurrence, meaning the joint appearance of words (cf. Blei et al. 2003, 996, note 1; Althage 2022, 267; see also Shadrova 2021). Within the context of humanities research, however, applying Topic Modeling is associated with two assumptions: First, that of topic coherence, which states that the terms assigned to a topic should have a thematic or conceptual relationship; and second, the assumption of stability of meaning, according to which a given topic, if assigned to different documents, should have the same meaning or relevance for all these documents (Schmidt 2012, 49).

4 At the same time, however, this also means that the type and scope of data preprocessing has a significant influence on the modeling result. Cf. section 4.2.

However, Topic Models do not *understand* the meaning and concepts that people associate with the words of a text, given the computer is "semantically blind" in this respect (Schwandt 2018, 108. 133). Accordingly, Benjamin Schmidt critically pointed out that topics are not inherently meaningful but become so through our interpretation (Schmidt 2012; see also Horstmann 2018, 10). David Blei noted that topics could resemble themes because words that often co-occur tend to be part of the same thematic field (Blei 2012a, 9). As such, claiming these topics as themes or discourses is based on the principle of distributional semantics (Piper 2018, 13; Schöch 2017, 14). Distributional semantics formalizes the assumption that the meaning of words arises from their co-occurrence frequency with other words in a particular context. The context can be a document, a paragraph, or even a single sentence. To interpret text data on a "semantic level", these frequency relationships between words are numerically represented by, for example, coordinates in a vector space, making them computationally processable (Turney & Pantel 2010; Blei 2012a, 9; Piper 2018, 13–18; see also Althage 2022, 266 f.).

Despite topics being frequently equated with *themes* or other semantic categories, they are not synonymous (Uglanova & Gius 2020, 72). This is also reflected in the fact that a topic model usually also includes word clusters – sometimes heavily weighted – that describe more general stylistic properties of a specific text type (*meta-topics,* see Fig. 2) or indicate heterogeneity in the text data, which is reflected, for example, in language-specific topics (cf. Fig. 3; on various topic forms, see Boyd-Graber et al. 2014, 234–237; Schöch 2017, 23–26; Althage 2022, 267–269). The latter, referred to as *Noisy Topics,* could serve as a starting point for further preprocessing of the text data (see section 4.2). In the context of the aforementioned assumptions, it should also be noted that a topic with the same weighting in two different documents can also have completely different emphases at the word level.[5] Topics are thus not independent epistemological units with a fixed semantic core, but rather a hermeneutic instrument (Rockwell & Sinclair 2016). Topics enable a structured approach to large amounts of text yet always require interpretation performed within the context of the underlying sources, whereby the assumptions of coherence and stability of meaning in particular should be examined.

5 See, e. g., the reviews at https://www.hsozkult.de/publicationreview/id/reb-25382 and https://www.hsozkult.de/publicationreview/id/reb-26856, which, with a distribution of 32 % each for topic 42 (see Fig. 1), reflect different conceptual dimensions of the topic. Both addresses were accessed on 18 June 2024.

Fig. 2 Examples of *Metatopics* from the book reviews of H-Soz-Kult

Fig. 3 Examples of *Noisy Topics* from the book reviews of H-Soz-Kult

4. Topic Modeling Workflow

The concrete application of Topic Modeling in a research context requires a carefully constructed, critically reflected, and documented workflow (see Fig. 4). The workflow includes the selection of a suitable method, the preparation of the text data *(preprocessing)* for the generation of Topic Models and the evaluation of the results, considering various configurations of preprocessing and Topic Modeling. Generally, this process is iterative where it is possible to go back and forth between the individual processing steps to optimize the modeling results regarding the research question. Once an appropriate Topic Model has been found, there are various visualization options for the results, from simple word lists and word clouds to bar graphs, line charts, or scatter plots to illustrate characteristics and developments, and heat maps for correlations or networks for relationships between the clusters. This section focuses on the selection of a suitable algorithm, the preprocessing of the data, and the evaluation of the results.

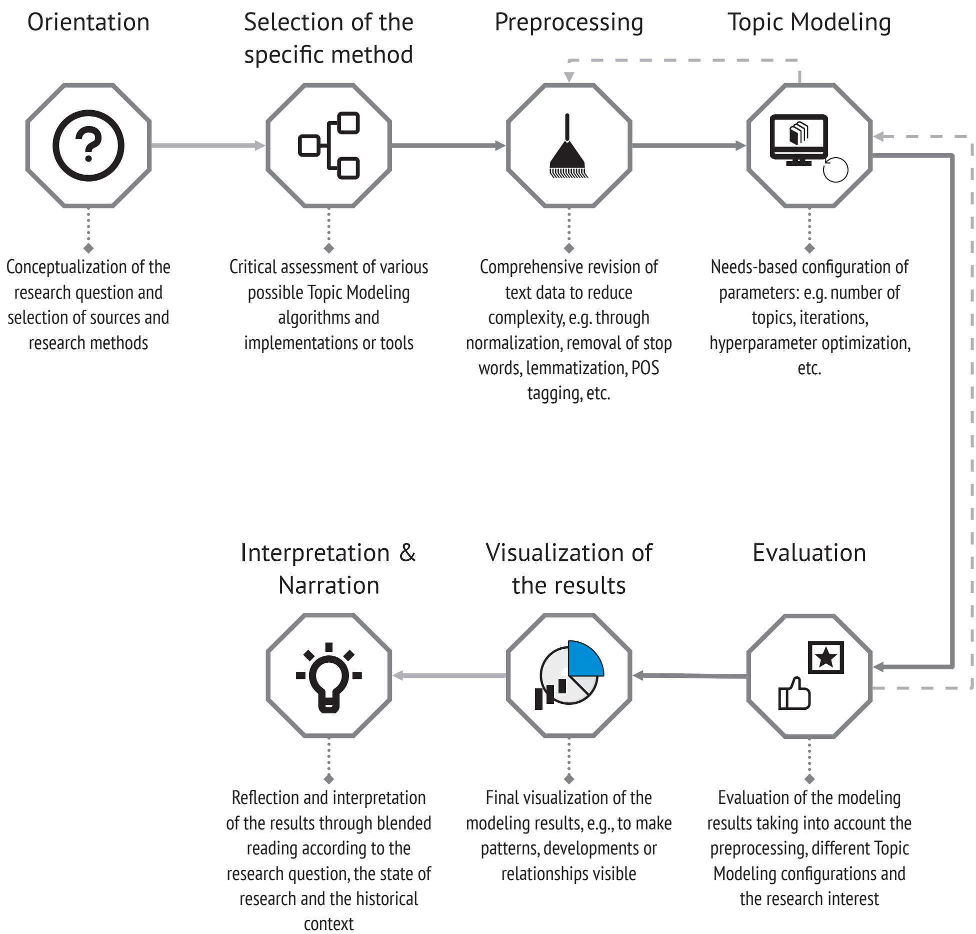

Fig. 4 Schematic Topic Modeling workflow

4.1 Selection of the Topic Modeling Method

At the beginning of the research project, it should be determined which algorithm in which implementation is suitable for a given research question (cf. Jelodar et al. 2019; Vayansky & Kumar 2020; Churchill & Singh 2022). This decision should be based on a comparison of various approaches and their respective results (Fig. 5 is an exemplary aid for decision-making). Various factors must be considered, e.g., the consistency of the theoretical-methodological assumptions of the procedure with one's own research objectives and the available configuration options (from the number of topics to hyperparameter optimization) and their effects on the output. The aim is to critically examine the potentials and limitations of the methods under consideration.

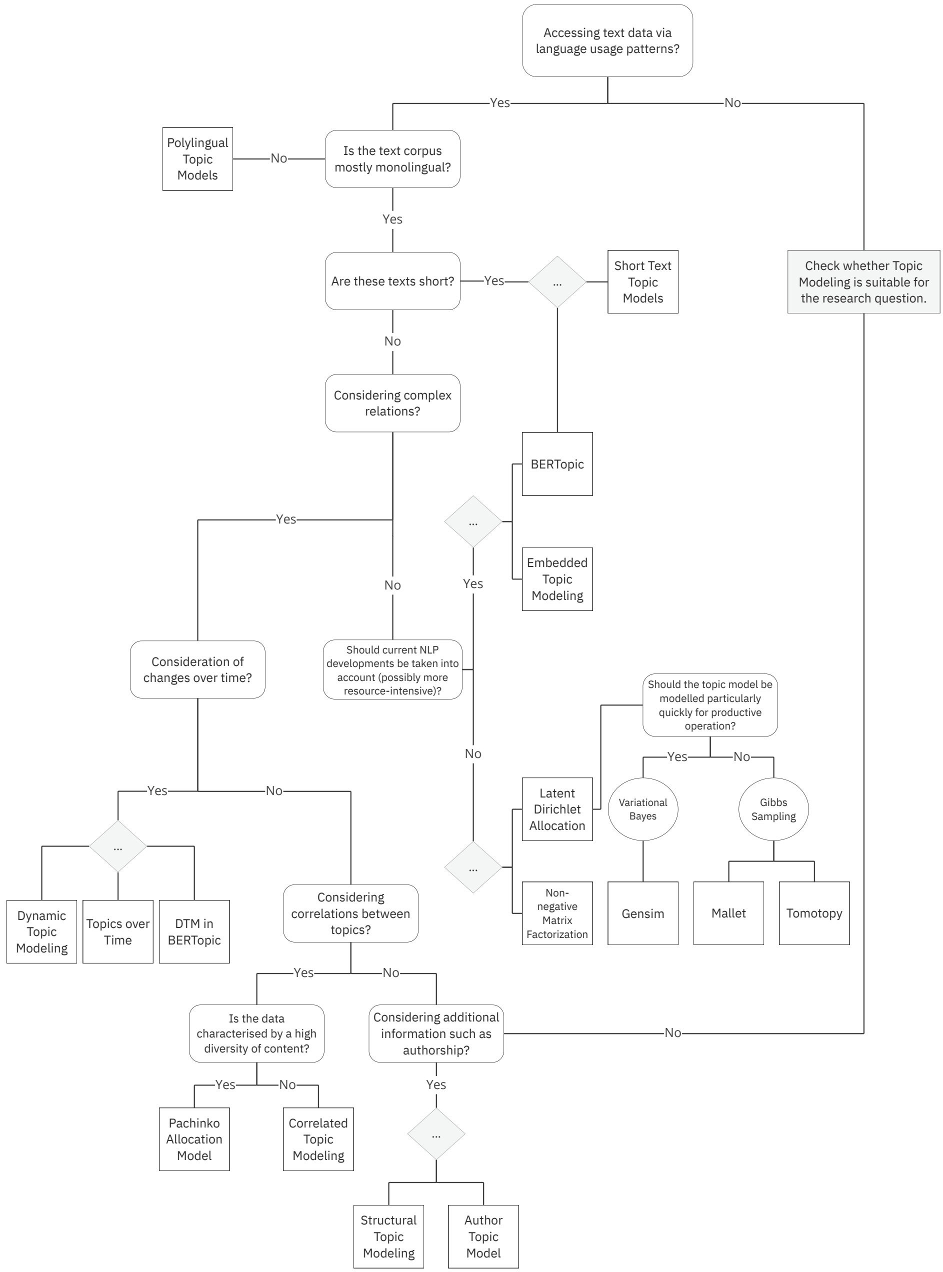

Fig. 5 Exemplary decision tree (building on Vayansky & Kumar 2020, esp. 14, Fig. 8; Churchill & Singh 2022; Jelodar et al. 2019); some key questions about the goal or the characteristics of the research object can help to select the appropriate method or tool.*

* On *Non-negative Matrix Factorization* see: Lee & Seung 1999; *Topics over Time:* Wang & McCallum 2006; *Pachinko Allocation Model:* Li & McCallum 2006; on *Embedded Topic Modeling* for example: Dieng et al. 2020.

Among the numerous options, LDA (Blei et al. 2003) has proven to be particularly popular in the Digital Humanities and is also implemented in numerous ready-to-use tools and programming libraries.[6] This method has been used successfully as a heuristic tool in a variety of disciplines to explore extensive text collections. In the Digital Humanities, the method is also used for historical studies that extend over longer periods of time (such as, Wehrheim et al. 2022; Snickars 2022; Grant et al. 2021). However, since LDA does not consider the temporal and relational dimension of the data or its contextuality in the modeling process, the information must be applied to the model retrospectively (Althage 2022). In contrast, methods like *Dynamic Topic Modeling* (DTM; Blei & Lafferty 2006; Grootendorst 2022; for an application of the method, cf. Guldi 2019) take the temporality of the topics into account during the modeling process, allowing them to show how, for example, discourses emerge, develop, and disappear over time.

Should the focus be on the relationships between different clusters rather than the temporal dimensions, there are suitable methods like *Correlated Topic Modeling* (CTM; Lafferty & Blei 2005; Blei & Lafferty 2007). CTM can show how different topics correlate with each other. *Structural Topic Modeling* (STM; Roberts et al. 2014; Küsters & Garrido 2020), on the other hand, enables one to model topics in relation to specific contextual information, which is particularly useful when investigating the influence of factors such as gender, social group affiliation, or genre on topic formation. The influence of authorship on the topic model can also be investigated using STM, but special *Author Topic Models* have also been developed (Rosen-Zvi et al. 2004).

The versatility of Topic Modeling is evident in its applicability to different types of text, from scientific articles to historical documents and social media posts. However, for shorter texts, for instance tweets or titles of works, specialized models such as *Short Text Topic Models* (Cheng et al. 2014; Zuo et al. 2016; Zhao et al. 2021) may be more suitable. For multilingual text collections, *Polylingual Topic Models* (Mimno et al. 2009) or *BERTopic* (Grootendorst 2022) can be used to identify thematic consistencies across different languages.

After providing an overview of various Topic Modeling methods and their application, the question arises as to how these models practically can be integrated into the research process. There are various options: for example, ready-to-use tools such as the *TopicsExplorer*[7] or *Topics* in *Voyant Tools*[8] can be used. Because the configu-

6 However, it should be noted that there are different implementations of LDA, which can generate different modeling results. For example, the implementations in *MALLET* and *Gensim* differ in terms of their inference algorithms for deriving the topics. The *Gensim* implementation is designed to handle very large amounts of data and focuses on performance; the results can therefore be less coherent. In contrast, *MALLET* requires more computational time to model the topics, but generally produces more coherent and robust modeling results – even with smaller text corpora. Cf. Althage 2022, 261–263; Hodel et al. 2022; Boyd-Graber et al. 2014, 231–233.

7 See https://dariah-de.github.io/TopicsExplorer (Accessed: 19 July 2024).

8 See http://voyant-tools.org (Accessed: 18 June 2024), cf. Rockwell & Sinclair 2016..

ration options have a substantial impact on the results, especially with this method, these software solutions should primarily be seen as an introductory aid to familiarize oneself with the modeling process. While the number of topics to be generated and their iterations can often be freely selected using these tools, more complex components such as hyperparameters, which influence the distribution profile of the topics (Wallach et al. 2009; Boyd-Graber et al. 2014, 233), are hidden in the *Black Box*. The possibilities for evaluating the modeling results or exporting them as reusable data are also limited. Furthermore, it should be noted that Topic Modeling, as we have seen, represents a whole range of algorithms that all pursue, in one way or another, the goal of grouping texts based on their patterns of language use in order to explore their thematic structure.

It is therefore advisable to use more complex solutions such as the widely used framework *MALLET,*[9] the *interactive Leipzig Corpus Miner* (iLCM),[10] or implementations of various algorithms in programming languages such as *Python* or *R*, enabling one to configure the procedures according to one's needs. Libraries in Python such as *Gensim,*[11] *Scikit-Learn,*[12] *Tomotopy,*[13] *BERTopic,*[14] or *OCTIS,*[15] for example, offer several solutions within a single environment.[16] Choosing the appropriate algorithm and implementation is only the first step in a complex process; as we will discuss in the next chapter, the careful preparation of the text corpus is essential for historical and stylistically diverse text data.

4.2 Preprocessing

Topic Modeling can, in principle, be applied to any text from any language. Although especially for historical research disciplines, one should consider that the methods presented above were usually developed and tested using text data corresponding to modern languages, which are more standardized than medieval or early modern texts. Additionally, literary texts with their numerous stylistic peculiarities can also pose a challenge in this context (see on this for example Uglanova & Gius 2020). LDA, e.g., was tested using, among others, English language news and scientific articles (Blei et al. 2003). As the method makes regularities visible, a corpus linguisti-

9 See http://mallet.cs.umass.edu (Accessed: 18 June 2024), cf. McCallum 2002.
10 See https://ilcm.informatik.uni-leipzig.de (Accessed: 18 June 2024), cf. Niekler et al. 2023.
11 See https://radimrehurek.com/gensim (Accessed: 18 June 2024), cf. Řehůřek & Sojka 2010.
12 See https://scikit-learn.org/stable/index.html (Accessed: 18 June 2024).
13 See https://bab2min.github.io/tomotopy (Accessed: 18 June 2024).
14 See https://maartengr.github.io/BERTopic/index.html (Accessed: 19 July 2024), cf. Grootendorst 2022.
15 See https://github.com/MIND-Lab/OCTIS (Accessed: 18 June 2024), cf. Terragni et al. 2021.
16 A look at the documentation of the libraries provides information about the configuration options and initial sample code. Usually, there are also numerous useful tutorials available online.

cally and orthographically homogeneous is more reliable to model than 17th century funeral writings, which were not yet subject to comparable written language rules and can have different spellings for the same concepts as well as numerous Latin remarks and quotations. The more complex and varied the sources, the less consistent and predictable the results of the modeling may potentially be.

Tab. 1 Exemplary comparison of a selection of topics before and after initial preprocessing (17th century funeral sermons, "AEDit Frühe Neuzeit")

Before Preprocessing	**After Preprocessing (Tokenization, Removal of Punctuation Marks and Numbers, Lemmatization, POS tagging, Lowercasing)**
und der die das zu mit auch er nicht den dem ist sie von ein wie des sich Gott daß	kind eltern job söhnlein kinderlein kindlein lieb gerecht taufe bräutigam töchterlein braut gerechtigkeit item de christus matt arm justitia himmlisch
Frau Kinder Mutter Adeligen Eltern Kind Adelige Gn. Edlen/E. J. Kinderlein liebes Weib Eltern/Rahel geborene Söhnlein Kindlein Job	frau lieb mutter adelig junker weib adelige witwe edl gn geboren rahel herz schmerz trost schwester gestreng kind kreuz augenlust
Prediger Lehrer & Kirchen Amt M. Zuhörer Stadt ad Anno D. Prediger/c. treuen Gemeine Schulen Fürstlichen Fürstl. treue	prediger kirche lehrer amt jahr zuhörer wort treu prophet groß lehre schule stadt apostel anno ehrwürdig predigt knecht fürstl mann
Dann dann wann wider sonder „ Vers deren lang dieselbige Leibs Tods Kapitel gern Edlen Arzt Sara Sohns seliger dieweil	christus arzt jesus arznei kreuz kapitel doktor luc matth apotheke christi wunde joh apotheker volk hiob medikus leiden heiland jude
Sie Er daß die der Ihr als eine Ihm von zu Frau Die Ich dem den GOTT sich Der Ihre	frau seele hoch himmel welt mutter tod haus freude träne auge herrlichkeit ps leben liebe braut ehre vater land tugend

To filter out the content characteristics of the texts (see Tab. 1), it is advisable to reduce the complexity of the text data by standardizing and normalizing the vocabulary; this work step is called *preprocessing* (cf. Maier et al. 2018, 97f. 100–102. 110). The type, scope, and sequence of the individual processing steps are significant and depend on the chosen method, as well as on the type of source to be processed and the respective research question. When selecting and arranging the individual processing steps, one has to consider the language and the degree of standardization and normalization. While the available resources for modern languages are increasing, the availability for historical languages is still limited, which can result in more complex preprocessing. Be it modern or historical texts, one must proceed carefully when preparing the data and documenting the individual decisions in this often iterative process to ensure the traceability of the procedure.

Tokenization is one of the mandatory preparatory steps of numerous text analysis methods. This involves breaking down the text into smaller units, also known as *tokens,* which can then be processed, counted, compared, and recombined. Tokenization is usually performed on words. While humans can intuitively recognize lexical units, this process must be explicitly formalized for the computer. Depending on the language, this computation poses its own challenges. The handling of multi-word units is particularly relevant, which are sometimes but not always identified by hyphens, as in the case of "Holy Spirit." Typically, the connection between the two words would be removed during tokenization ("Holy," "Spirit"), so that the individual word components are processed independently of each other. Therefore, it is essential to reflect on what should be considered as the unit of investigation in the respective research context. For such *Natural Language Processing* tasks, numerous established tools already exist.[17] Moreover, the modeling of *bigrams* (word pairs) and *trigrams* (word triples) can also assist in reassembling phrases composed of co-occurring terms into a single token, thereby partially preserving the local reference.

There is no prescribed path for further preprocessing; however, several steps have been established that can be used modularly and tailored to the data corpus, depending on the intended purpose (see Fig. 6). One such step is *segmentation.* Particularly when dealing with corpora made up of very extensive individual documents (e.g., a corpus of books), the documents can be broken down into smaller units (e.g., at chapter or paragraph level). The removal of punctuation and the so-called stop words has also become common practice (see e.g. Schofield et al. 2017 for a critical perspective on the removal of stop words). The stop words are function words like "the," "in," "at," etc., which, as they occur very frequently in texts, would dominate the topics as statistically very prominent characteristics of texts (see Tab. 1). Even if these words are of central grammatical importance, they do not necessarily belong to the words that carry meaning and could make working with the Topic Model more difficult. Depending on the language, there are various stop word lists that can be reused through programming libraries, such as NLTK, and extended manually with additional terms specific to the corpus. The contents of these lists should be checked to ensure that relevant words to one's project are not removed.

Methods like *TF-IDF* or *Part-of-Speech Tagging* (POS tagging) offer a more systematic approach to selecting relevant and meaningful words than stop word lists. With TF-IDF, one can identify tokens that are characteristic of a particular document or group of documents and, in contrast, give less weight to words that occur particularly frequently in many documents (such as function words, but also other corpus-specific terms) and filter them out accordingly (Klinke 2017, 274 f.). POS tagging, on the other hand, automatically determines the word types of the lexical units. In this way, spe-

17 In *Python,* e.g., see NLTK. Natural Language Toolkit, URL: https://www.nltk.org; spaCy. Industrial-Strength Natural Language Processing, URL: https://spacy.io. Both addresses were accessed on 18 June 2024.

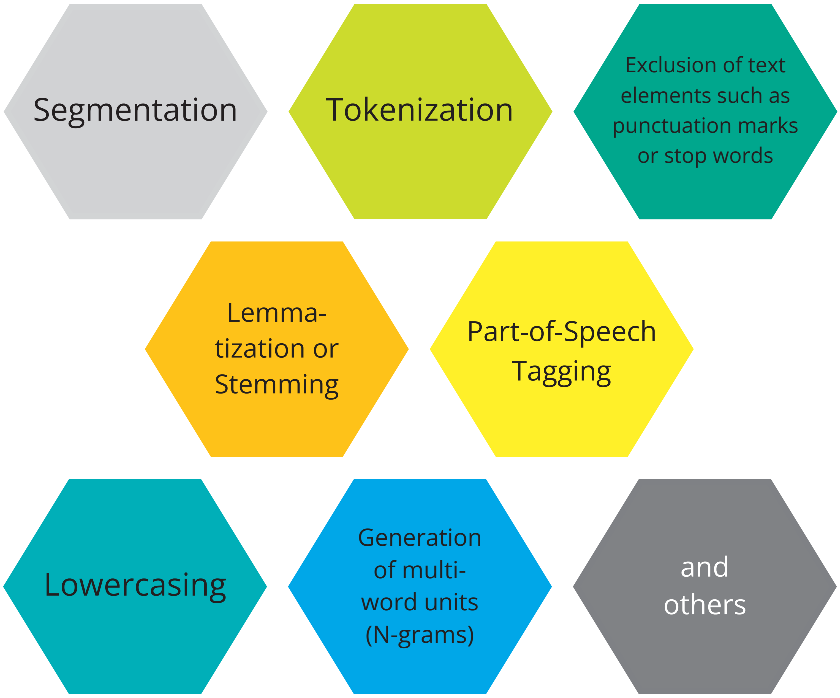

Fig. 6 The preparation of text data can be composed of various modular steps depending on the characteristics and the quality of the data as well as the research objectives.

cific word types can be selected for analysis that are assumed to have a meaningful function in texts (e.g., nouns, verbs, or adjectives; see for example, Schöch 2017, 17).

To minimize the variations in word forms and enable the modeling of more coherent topics, as well as to facilitate data processing, in addition to *lowercasing* (i.e., the lowercasing of all tokens), *lemmatization* has proven to be efficient. This approach reduces inflected single word forms to their base form (holier → holy, went → go). *Stemming* is also not uncommon, especially in English contexts. The individual words are reduced to their stem or root by truncating the word endings (e.g., Protestant → Protest, Protestantism → Protestant[18]). Stemming results in tokens that do not necessarily reflect a valid lexical entry in a language and can therefore be much more difficult to interpret (Schofield & Mimno 2016). If the sole aim is to index a corpus efficiently, then the latter approach might be appropriate; however, lemmatization is the preferred option for research projects aimed at interpretation.

18 Editor's note: A different example was chosen in the original text, which does not work so well in English: Christian, Christ → Christ, Christianity → Christian.

Each step of processing has a direct effect on the respective modeling result and thus on what is intended to be interpreted.[19] The procedure should therefore not only be documented, but also integrated into the evaluation of the Topic Models, taking into account the research question and knowledge objectives.

4.3 Evaluation

Evaluating the Topic Models is essential to ensure the quality and relevance of the generated clusters. This is especially advisable as there is a risk of succumbing to *confirmation bias* with such methods, i. e. processing the modeling and the resulting data until a desired or expected outcome is achieved (Shadrova 2021, 5, 16 f.). Although, as Maria Antoniak pointed out, Topic Modeling is not about generating the one "correct" perspective of the text corpus, but about supporting a qualitative investigation by discovering one of many possible "interpretative lenses" through which sources can be understood (Antoniak 2022), it is advisable to also include mathematical evaluation metrics in addition to the qualitative examination of the topics in terms of their interpretability and representativeness (a good introduction is provided by Boyd-Graber et al. 2014, 233 f. 237–243; Churchill & Singh 2022, 5–9).

For qualitative evaluation, *Blended Reading* can be applied as an evaluation mode (Stulpe & Lemke 2016). This approach combines the results of the machine learning process *(Distant Reading)* with human reading and interpretation *(Close Reading).* By reading representative documents or text passages for each topic, one can evaluate their interpretability and representativeness. Or by comparing the most important words and phrases assigned to the topics, the granularity of the model can be determined.[20] It can also be useful to check whether there is a Ground Truth for the text corpus or whether it is feasible to create one. This means, for example, an already existing manual classification that the model can be compared against.[21] In this case, metrics such as *accuracy, precision, recall,* and *F-score* can be used to evaluate the performance of the model (Churchill & Singh 2022, 5–9; Klinke 2017, 269 f.).

Since there is usually no such Ground Truth when using Topic Modeling, a number of other metrics have been established that can be used to evaluate the modeling results (Churchill & Singh 2022, 6–8; Boyd-Graber et al. 2014, 233 f. 237–243), including:

19 Newer methods such as *BERTopic*, promise to be able to dispense with preprocessing by using the latest language models. However, it remains to be seen how well this works for historical languages.

20 The number of topics has an impact on the granularity of the model. Too high a number potentially leads to overlapping, redundant clusters, while too low a number results in clusters that are too heterogeneous, see also Schöch 2017, 20, note 7.

21 In case of H-Soz-Kult, for example, there is a manual classification according to themes, regions, and epochs, which provides a good orientation for the evaluation of the modeling results.

- Coherence measures can be used, for example, to measure how well the (top) words assigned to the topics fit together: the higher the coherence value, the more semantically coherent the topics and thus interpretable the word clusters are in theory. Tools such as *Gensim* in *Python* offer functions for calculating coherence, which can also be used to determine which number of topics is appropriate for a given research topic.[22]
- Perplexity, on the other hand, can be used to assess how well the topic model can predict new, unseen documents. A lower perplexity value is typically better, but this value alone is often not sufficient to assess the quality of a model.
- The exclusivity or uniqueness of the (top) words assigned to the topics can also be measured for the respective topics in order to assess the distinctiveness of the word clusters.

The above are just a few of the available options for evaluating topic models. Since these evaluation metrics do not always correlate positively with human assessments of the modeling results (e.g. Hoyle et al. 2021; Uglanova & Gius 2020), they should always be used in addition to the qualitative manual interpretation by the researchers with their domain knowledge and with due consideration of the specific research question and the characteristics of the text corpus.

5. Concluding Remarks

In conclusion, it can be stated that Topic Modeling can be a valuable addition to the methodological landscape of theology. As a “statistical lens” that formalizes the knowledge, theories, and assumptions of theology (after Blei 2012a, 8), it can provide new data-driven perspectives on sources and research debates. Although Topic Models do not, by themselves, provide conclusions to specific research questions, they can be used to explore hypotheses and test them against the individual sources. Topic Modeling does not replace textual hermeneutic approaches, but rather expands the study of sources with an additional set of tools. Thus, it bridges the gap between traditional hermeneutic approaches and modern, data-based methods. It enriches the analytical repertoire of the humanities and creates new avenues for the systematic and critical examination of extensive text corpora. It is to be hoped that this approach will serve as a starting point for further investigations and discussions within theology.

22 With *Hierarchical Dirichlet Process* (HDP), an extension of LDA, a method was developed making it possible to derive the number of topics from the corpus data (Teh et al. 2006).

References

Althage, M. (2022). Potenziale und Grenzen der Topic-Modellierung mit Latent Dirichlet Allocation für die Digital History. In K.D. Döring, S. Haas, M. König & J. Wettlaufer (Eds.), *Digital History. Konzepte, Methoden und Kritiken Digitaler Geschichtswissenschaft* (pp. 255–277). Berlin/Boston: De Gruyter Oldenbourg [= *Studies in Digital History and Hermeneutics,* 6]. DOI: https://doi.org/10.1515/9783110757101-014 (Accessed: 18 June 2024).

Ead. (2023). Digitale Methoden kritisch reflektieren. Die Erweiterung des Werkzeugkastens der Historiker:innen. In P. Trilcke, A. Busch & P. Helling (Eds.), *DHd 2023. Open Humanities Open Culture. 9. Tagung des Verbands Digital Humanities im deutschsprachigen Raum.* Trier/Luxemburg: Zenodo. DOI: https://doi.org/10.5281/zenodo.7711522 (Accessed: 18 June 2024).

Antoniak, M. (2022). Topic Modeling for the People. In *Blog by M. Antoniak.* URL: https://maria-antoniak.github.io//2022/07/27/topic-modeling-for-the-people.html (Accessed: 18 June 2024).

Blei, D.M. (2012a). Topic Modeling and Digital Humanities, *Journal of Digital Humanities,* 2(1), 8–11. URL: http://journalofdigitalhumanities.org/2-1/topic-modeling-and-digital-humanities-by-david-m-blei (Accessed: 18 June 2024).

Id. (2012b). Probabilistic topic models, *Communications of the ACM,* 55(4), 77–84. DOI: https://doi.org/10.1145/2133806.2133826 (Accessed: 18 June 2024).

Id., & Lafferty, J.D. (2006). Dynamic Topic Models. In *Proceedings of the 23rd International Conference on Machine Learning – ICML '06* (pp. 113–120). New York: ACM Press. DOI: https://doi.org/10.1145/1143844.1143859 (Accessed: 18 June 2024).

Eid. (2007). A correlated topic model of Science, *The Annals of Applied Statistics* 1(1), 17–35. DOI: https://doi.org/10.1214/07-AOAS114 (Accessed: 18 June 2024).

Blei, D.M., Ng, A.Y., & Jordan, M.I. (2003). Latent Dirichlet Allocation, *Journal of Machine Learning Research,* 3, 993–1022. DOI: https://dl.acm.org/doi/10.5555/944919.944937 (Accessed: 21 July 2024).

Boyd-Graber, J., Mimno, D., & Newman, D. (2014). Care and Feeding of Topic Models. Problems, Diagnostics, and Improvements. In E.M. Airoldi, D.M. Blei, E.A. Erosheva & S.E. Fienberg (Eds.), *Handbook of Mixed Membership Models and Their Applications* (pp. 225–254). London/New York: Chapman and Hall/CRC. DOI: https://doi.org/10.1201/b17520-21 (Accessed: 18 June 2024).

Brett, M.R. (2012). Topic Modeling. A Basic Introduction, *Journal of Digital Humanities,* 2(1), 12–16. URL: http://journalofdigitalhumanities.org/2-1/topic-modeling-a-basic-introduction-by-megan-r-brett (Accessed: 18 June 2024).

Bunout, E., & von Lange, M. (2019). Nibbling at Text. Identifying Discourses on Europe in a Large Collection of Historical Newspapers Using Topic Modelling. In *C2DH | Luxembourg Centre for Contemporary and Digital History.* URL: https://www.c2dh.uni.lu/thinkering/nibbling-text-identifying-discourses-europe-large-collection-historical-newspapers-using (Accessed: 18 June 2024).

Cheng, X., Yan, X., Lan, Y., & Guo, J. (2014). BTM. Topic Modeling over Short Texts, *IEEE Transactions on Knowledge and Data Engineering,* 26(12), 2928–2941. DOI: https://doi.org/10.1109/TKDE.2014.2313872 (Accessed: 18 June 2024).
Churchill, R., & Singh, L. (2022). The Evolution of Topic Modeling, *ACM Computing Surveys,* 54(10s), 1–35. DOI: https://doi.org/10.1145/3507900 (Accessed: 18 June 2024).
Crane, G. (2006). What Do You Do with a Million Books?, *D-Lib Magazine,* 12(3), no pag. DOI: https://doi.org/10.1045/march2006-crane (Accessed: 18 June 2024).
Dieng, A.B., Ruiz, F.J.R., & Blei, D.M. (2020). Topic Modeling in Embedding Spaces, *Transactions of the Association for Computational Linguistics,* 8, 439–453. DOI: https://doi.org/10.1162/tacl_a_00325 (Accessed: 21 July 2024).
Graham, Sh., Milligan, I., & Weingart, S. (2016). *Exploring Big Historical Data. The Historian's Macroscope.* London: Imperial College Press.
Grant, Ph., Sebastian, R., Allassonnière-Tang, M., & Cosemans, S. (2021). Topic Modelling on Archive Documents from the 1970s. Global Policies on Refugees, *Digital Scholarship in the Humanities,* 36(4), 886–904. DOI: https://doi.org/10.1093/llc/fqab018 (Accessed: 18 June 2024).
Graves, M. (2022). Computational Topic Models for Theological Investigations, *Theology and Science,* 20(1), 69–84. DOI: https://doi.org/10.1080/14746700.2021.2012922 (Accessed: 18 June 2024).
Grootendorst, M. (2022). BERTopic. Neural topic modeling with a class-based TF-IDF procedure. Online: *arXiv.* DOI: https://doi.org/10.48550/arXiv.2203.05794 (Accessed: 18 June 2024).
Guldi, J. (2019). Parliament's Debates about Infrastructure. An Exercise in Using Dynamic Topic Models to Synthesize Historical Change, *Technology and Culture,* 60(1), 1–33. DOI: https://doi.org/10.1353/tech.2019.0000 (Accessed: 18 June 2024).
Hodel, T., Möbus, D., & Serif, I. (2022). Von Inferenzen und Differenzen. Ein Vergleich von Topic-Modeling-Engines auf Grundlage historischer Korpora. In S. Gerlek, S. Kissler, Th. Mämecke, & D. Möbus (Eds.), *Von Menschen und Maschinen. Mensch-Maschine-Interaktionen in digitalen Kulturen* (pp. 185–209). Hagen: Hagen University Press. DOI: https://doi.org/10.57813/20220623-153139-0 (Accessed: 21 July 2024)
Horstmann, J. (2018). Topic Modeling, *ForText. Literatur Digital Erforschen,* 1–16. URL: https://fortext.net/routinen/methoden/topic-modeling (Accessed: 18 June 2024).
Hoyle, A., Goel, P., Peskov, D., Hian-Cheong, A., Boyd-Graber, J., & Resnik, Ph. (2021). Is Automated Topic Model Evaluation Broken? The Incoherence of Coherence. In *Advances in Neural Information Processing Systems,* 34 (pp. 2018–2033). URL: https://proceedings.neurips.cc/paper_files/paper/2021/hash/0f83556a305d789b1d71815e8ea4f4b0-Abstract.html (Accessed: 18 June 2024).
Jelodar, H., Wang, Y., Yuan, Ch., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent Dirichlet Allocation (LDA) and Topic Modeling. Models, Applications, a

Survey, *Multimedia Tools and Applications,* 78(11), 15169–15211. DOI: https://doi.org/10.1007/s11042-018-6894-4 (Accessed: 18 June 2024).
Jockers, M.L. (2013). *Macroanalysis. Digital Methods and Literary History.* Urbana/Chicago/Springfield: University of Illinois Press.
Klinke, H. (2017). Information Retrieval. In F. Jannidis, H. Kohle, & M. Rehbein (Eds.), *Digital Humanities. Eine Einführung* (pp. 268–278). Stuttgart: J.B. Metzler. DOI: https://doi.org/10.1007/978-3-476-05446-3_19 (Accessed: 18 June 2024).
Küsters, A., & Garrido, E. (2020). Mining PIGS. A Structural Topic Model Analysis of Southern Europe Based on the German Newspaper Die Zeit (1946–2009), *Journal of Contemporary European Studies,* 28(4), 477–493. DOI: https://doi.org/10.1080/14782804.2020.1784112 (Accessed: 18 June 2024).
Lafferty, J.D., & Blei, D.M. (2005). Correlated Topic Models. In *Advances in Neural Information Processing Systems,* 18 (pp. 147–154). URL: https://proceedings.neurips.cc/paper_files/paper/2005/hash/9e82757e9a1c12cb710ad680db11f6f1-Abstract.html (Accessed: 18 June 2024).
Lee, D.D., & Seung, H.S. (1999). Learning the parts of objects by non-negative matrix factorization, *Nature,* 401, 788–791. DOI: https://doi.org/10.1038/44565 (Accessed: 21 July 2024).
Li, W., & McCallum, A. (2006). Pachinko Allocation. DAG-Structured Mixture Models of Topic Correlations. In *Proceedings of the 23rd International Conference on Machine Learning* (pp. 577–584). New York: Association for Computing Machinery. DOI: https://doi.org/10.1145/1143844.1143917 (Accessed: 21 July 2024).
Luhmann, J., & Burghardt, M. (2021). Digital Humanities. A Discipline in Its Own Right? An Analysis of the Role and Position of Digital Humanities in the Academic Landscape, *Journal of the Association for Information Science and Technology,* 73(2), 148–171. DOI: https://doi.org/10.1002/asi.24533 (Accessed: 18 June 2024).
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. DOI: https://doi.org/10.1080/19312458.2018.1430754 (Accessed: 18 June 2024).
McCallum, A.K. (2002). MALLET. A Machine Learning for Language Toolkit. URL: http://mallet.cs.umass.edu (Accessed: 18 June 2024).
Mimno, D. (2012). Computational Historiography. Data Mining in a Century of Classics Journals, *Journal on Computing and Cultural Heritage,* 5(1), 1–19. DOI: https://doi.org/10.1145/2160165.2160168 (Accessed: 18 June 2024).
Id., Wallach, H.M., Naradowsky, J., Smith, D.A., & McCallum, A. (2009). Polylingual Topic Models. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing* (pp. 880–889). Singapore: Association for Computational Linguistics. URL: https://aclanthology.org/D09-1092 (Accessed: 21 July 2024)

Niekler, A., Kahmann, Ch., Burghardt, M., & Heyer, G. (2023). The Interactive Leipzig Corpus Miner. An Extensible and Adaptable Text Analysis Tool for Content Analysis. *Publizistik,* 68, 325–354. DOI: https://doi.org/10.1007/s11616-023-00809-4 (Accessed: 18 June 2024).
Nunn, Ch.A. (2022). Das Thema patristischer Ethik. Versuch einer Annäherung durch Distanz am Beispiel der Briefe des Augustinus von Hippo, *Journal of Ethics in Antiquity and Christianity,* 4, 31–51. DOI: https://doi.org/10.25784/jeac.v4i0.1011 (Accessed: 18 June 2024).
Piper, A. (2018). *Enumerations. Data and Literary Study.* Chicago/London: The University of Chicago Press.
Řehůřek, R., & Sojka, P. (2010). Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks* (pp. 46–50). Valletta, Malta: ELRA.
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. DOI: https://doi.org/10.1111/ajps.12103 (Accessed: 18 June 2024).
Rockwell, G., & Sinclair, S. (2016). *Hermeneutica. Computer-Assisted Interpretation in the Humanities.* Cambridge, MA/London: The MIT Press.
Rosen-Zvi, M., Griffiths, Th., Steyvers, M., & Smyth, P. (2004). The Author-Topic Model for Authors and Documents. In *UAI '04. Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence* (pp. 487–494). Arlington: AUAI Press. DOI: https://dl.acm.org/doi/10.5555/1036843.1036902 (Accessed: 21 July 2024).
Schmidt, B.M. (2012). Words Alone. Dismantling Topic Models in the Humanities, *Journal of Digital Humanities,* 2(1), 49–65. URL: http://journalofdigitalhumanities.org/2-1/words-alone-by-benjamin-m-schmidt (Accessed: 18 June 2024).
Schöch, Ch. (2017). Topic Modeling Genre. An Exploration of French Classical and Enlightenment Drama, *Digital Humanities Quarterly,* 11(2), 1–53. URL: https://www.digitalhumanities.org/dhq/vol/11/2/000291/000291.html (Accessed: 19 July 2024).
Schofield, A., Magnusson, M., & Mimno, D. (2017). Pulling Out the Stops. Rethinking Stopword Removal for Topic Models. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, 2. Short Papers* (pp. 432–436). URL: https://www.aclweb.org/anthology/E17-2069 (Accessed: 18 June 2024).
Schofield, A., & Mimno, D. (2016). Comparing Apples to Apple. The Effects of Stemmers on Topic Models, *Transactions of the Association for Computational Linguistics,* 4, 287–300. DOI: https://doi.org/10.1162/tacl_a_00099 (Accessed: 18 June 2024).
Schwandt, S. (2018). Digitale Methoden für die Historische Semantik. Auf den Spuren von Begriffen in digitalen Korpora, *Geschichte und Gesellschaft. Zeitschrift für historische Sozialwissenschaft,* 44(1), 107–134. DOI: https://doi.org/10.13109/gege.2018.44.1.107 (Accessed: 18 June 2024).

Shadrova, A. (2021). Topic Models Do Not Model Topics. Epistemological Remarks and Steps towards Best Practices, *Journal of Data Mining & Digital Humanities,* 1–28. DOI: https://doi.org/10.46298/jdmdh.7595 (Accessed: 18 June 2024).
Simmler, S., Vitt, Th., & Pielström, S. (2019). Topic Modeling with Interactive Visualizations in a GUI Tool. In *Proceedings of the Digital Humanities Conference.* Utrecht. Tool: https://dariah-de.github.io/TopicsExplorer (Accessed: 18 June 2024).
Snickars, P. (2022). Modeling Media History. On Topic Models of Swedish Media Politics 1945–1989, *Media History,* 28(3), 403–424. DOI: https://doi.org/10.1080/13688804.2022.2079484 (Accessed: 18 June 2024).
Stulpe, A., & Lemke, M. (2016). Blended Reading. Theoretische und praktische Dimensionen der Analyse von Text und sozialer Wirklichkeit im Zeitalter der Digitalisierung. In M. Lemke & G. Wiedemann (Eds.), *Text Mining in den Sozialwissenschaften. Grundlagen und Anwendungen zwischen qualitativer und quantitativer Diskursanalyse* (pp. 17–61). Wiesbaden: Springer VS. DOI: https://doi.org/10.1007/978-3-658-07224-7_2 (Accessed: 21 July 2024).
Teh, Y.W., Jordan, M.I., Beal, M.J., & Blei, D.M. (2006). Hierarchical Dirichlet Processes, *Journal of the American Statistical Association,* 101(476), 1566–1581. DOI: https://doi.org/10.1198/016214506000000302 (Accessed: 18 June 2024).
Terragni, S., Fersini, E., Galuzzi, B.G., Tropeano, P., & Candelieri, A. (2021). OCTIS. Comparing and Optimizing Topic models is Simple! In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics. System Demonstrations* (pp. 263–270). DOI: https://doi.org/10.18653/v1/2021.eacl-demos.31 (Accessed: 18 June 2024).
Turney, P.D., & Pantel, P. (2010). From Frequency to Meaning. Vector Space Models of Semantics, *Journal of Artificial Intelligence Research,* 37(1), 141–188. DOI: https://jair.org/index.php/jair/article/view/10640 (Accessed: 18 June 2024).
Uglanova, I., & Gius, E. (2020). The Order of Things. A Study on Topic Modelling of Literary Texts. In *Proceedings of the Workshop on Computational Humanities Research* (CHR 2020) (pp. 57–76). URL: https://ceur-ws.org/Vol-2723/long7.pdf (Accessed: 18 June 2024).
Vayansky, I., & Kumar, S.A.P. (2020). A Review of Topic Modeling Methods, *Information Systems,* 94(101582), 1–15. DOI: https://doi.org/10.1016/j.is.2020.101582 (Accessed: 18 June 2024).
Völkl, Y., Sarić, S., & Scholger, M. (2022). Topic Modeling for the Identification of Gender-Specific Discourse. Virtues and Vices in French and Spanish 18th Century Periodicals, *Journal of Computational Literary Studies,* 1(1), 1–27. DOI: https://doi.org/10.48694/jcls.108 (Accessed: 18 June 2024).
Wallach, H.M., Mimno, D., & McCallum, A. (2009). Rethinking LDA. Why priors matter. In *Proceedings of the 22nd International Conference on Neural Information Processing Systems* (pp. 1973–1981). New York: Curran Associates Inc. DOI: https://dl.acm.org/doi/10.5555/2984093.2984314 (Accessed: 21 July 2024).

Wang, X., & McCallum, A. (2006). Topics over Time. A Non-Markov Continuous-Time Model of Topical Trends. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 424–433). New York: Association for Computing Machinery. DOI: https://doi.org/10.1145/1150402.1150450 (Accessed: 21 July 2024).

Wehrheim, L. (2019). Economic History Goes Digital. Topic Modeling the Journal of Economic History, *Cliometrica,* 13(1), 83–125. DOI: https://doi.org/10.1007/s11698-018-0171-7 (Accessed: 18 June 2024).

Id., Spoerer, M., & Jopp, T.A. (2022). Turn, Turn, Turn. A Digital History of German Historiography, 1950–2019, *The Journal of Interdisciplinary History,* 53(3), 471–507. DOI: https://doi.org/10.1162/jinh_a_01871 (Accessed: 18 June 2024).

Zhao, X., Wang, D., Zhao, Z., Liu, W., Lu, Ch., & Zhuang, F. (2021). A Neural Topic Model with Word Vectors and Entity Vectors for Short Texts, *Information Processing & Management,* 58(2), 1–11. DOI: https://doi.org/10.1016/j.ipm.2020.102455 (Accessed: 18 June 2024).

Zuo, Y., Wu, J., Zhang, H., Lin, H., Wang, F., Xu, K., & Xiong, H. (2016). Topic Modeling of Short Texts. A Pseudo-Document View. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 2105–2114). New York: Association for Computing Machinery. DOI: https://doi.org/10.1145/2939672.2939880 (Accessed: 18 June 2024).

Figure Credits

All figures were created by the author herself. Fig. 1–3 were generated with the Gensim wrapper for *MALLET* and visualised with the *Python* library *WordCloud*[23]. Fig. 4–6 were created with *Miro*[24].

Fig. 5f. are based on Althage (2023). All other figures were first published here.

23 See https://amueller.github.io/word_cloud (Accessed: 18 June 2024).
24 See https://miro.com/de (Accessed: 18 June 2024).