

Validating Topic Modeling as a Method of Analyzing Sujet and Theme

Julian Schröter¹ 
Keli Du² 

1. Lehrstuhl für Computerphilologie und Neuere Deutsche Literaturgeschichte, University of Würzburg , Würzburg, Germany.
2. Trier Center for Digital Humanities, University of Trier , Trier, Germany.

Citation

Julian Schröter and Keli Du (2022). "Validating Topic Modeling as a Method of Analyzing Sujet and Theme". In: *Journal of Computational Literary Studies* 1 (1). [10.48694/jcls.91](https://doi.org/10.48694/jcls.91)

Date published 2022-12-06

Date accepted 2022-03-29

Date received 2021-12-22

Keywords

sujet, theme, validation, topic modeling, content

License

CC BY 4.0 

Reviewers

Borja Navarro Colorado 

Christof Schöch 

Note

This paper has passed through the conference track of JCLS.

In addition to being peer reviewed, it was presented and discussed at the 1st Annual Conference of Computational Literary Studies 2022 at the Technical University of Darmstadt.

Abstract. In Computational Literary Studies (CLS), several procedures for thematic analysis have been adapted from NLP and Computer Science. Among these procedures, topic modeling is the most prominent and popular technique. We maintain, however, that this procedure is used only in the context of exploration up to date, but not in the context of justification. When we seek to prove assumptions concerning the correlation between genres, methods of computational text analysis have to be set up in research environments of justification, i.e. in environments of hypothesis testing. We provide a holistic model of validation and conceptual disambiguation of the notion of aboutness as sujet, fabula, and theme, and discuss essential methodological requirements for hypothesis-based analysis. As we maintain that validation has to be performed for individual tasks respectively, we shall perform empirical validation of topic modeling based on a new corpus of German novellas and comprehensive annotations and draw hypothetical generalizations on the applicability of topic modeling for analyzing aboutness in the domain of narrative fiction.

1. Introduction

Determining what literary texts are about is an essential part of interpreting literary texts and is also fundamental to investigating literary history. In Jockers (2013), which has been one of the most controversially received monographs in the last decade in computational literary studies (CLS), Jockers starts with a comprehensive and pretheoretical notion of theme, which is subsequently explored using topic modeling. Topic modeling is currently the most prominent tool for investigating aspects of aboutness in CLS. As it is based on unsupervised machine learning, topic modeling does not depend on our assumptions with regard to themes in texts. Hence, topic modeling has become a popular tool for exploring corpora. In several contexts, this tool has also been used in classification tasks for testing concrete hypotheses on genres or other text categories (e.g. Schöch 2017). The central claim of our paper is that topic modeling is still lacking justification to be used for hypothesis-driven research on specific aboutness claims in the domain of literary studies. Although this criticism on topic modeling is not new (e.g. Da 2019; Shadrova 2021), it has not yet been taken as a reason to overcome the desideratum. The task of this paper is to elaborate on this thesis and to prepare the methodological framework for solving this desideratum.

This desideratum affects the specific kind of interpretation that is at work when a concrete topic, which consists of a list of weighted words, is interpreted as, for example, a topic of ‘female fashion’ in Jockers and Mimno (2013), or of ‘love as a challenge and a reward’ in Schöch (2017). The use of topic modeling relies – at least implicitly – on the following three axioms in order to interpret lists of weighted words as genuine representations of aboutness:

1) A pre-theoretical notion has to be introduced to denote what topic modeling is expected to reveal in terms of humanities research. Our initial observation that Jockers (2013) starts from a general notion of ‘theme’ can be reversed. Theme is commonly considered to be the qualitative correlate of computationally generated topics. This holds also for Blei (2012), Jockers and Mimno (2013), Schöch (2017), and Weitin and Herget (2017). Hence, we take the linkage between the notion of theme and topic modeling to be the current state in CLS.

2) A specific theory of the structure of topics has to be developed. The formalized concept of topic in topic modeling can be outlined as follows: the core of topic modeling, Latent Dirichlet Allocation (LDA), comes from computational linguistics. It is a generative model and describes a fictional process in which a document is generated. It is based on the assumption that a text is a mixture of different topics with different probabilities, where each topic represents a probability distribution over a fixed set of words. A word can belong to one or several topics with certain probabilities. To generate a document, a probability distribution over topics is chosen randomly. Then, a topic is randomly chosen from all the topics and a word is randomly assigned to it. Thus, a single word of the document is determined. This process is then repeated until the document is finally generated (Blei 2012). LDA topic modeling in practice can then be understood as the inverse of the above described generative process. Given a text collection, the unseen topic-word distribution and the topic-document distribution are to be inferred by topic modeling (Blei and Lafferty 2009). There is often a semantic relationship between words that occur together in texts. These words are more likely to be grouped into one topic through topic modeling. Therefore, topics, which have in effect the form of lists of weighted words, are supposed to be interpretable as themes and to reflect the hidden content structure of the text collection.

3) A general theory is needed that justifies that themes are properly represented as lists of weighted words (topics), whose distribution in the text is similar. The best candidate of such a general theory seems to be distributional semantics, which holds that meaning consists of distributions of words (Evert 2005; Firth 1957; Harris 1954).

Based on these three steps, topic modeling is expected to return representations of ‘theme’ in a genuine sense of aboutness. However, our central claim that topic modeling lacks justification so far entails that topic modeling does not represent the genuine sense of aboutness in literary studies. In other words, the predicate “interpret a topic as a topic of...” is commonly used only in a loose sense, which means that the reader is reminded of a specific aboutness claim when reading a topic and expresses a subjective impression. If we want to use topic modeling for hypothesis-driven research on specific aboutness claims, the predicate should be used in a stricter sense that treats topics as an exact representation of specific aboutness claims. Section 2 of our paper elaborates and justifies our central claim. If we assume, for the moment, that our claim is correct, then

topic modeling is, at best, an approximation to aboutness under certain conditions. It is an approximation to aboutness if it can be substantiated with a more refined validation strategy. In general, the call for more validation is characteristic of CLS (Hammond 2017; Piper 2015; Swafford 2015). Such call for validation points to a methodological gap that arises when methods from domains such as statistics or computer linguistics are transferred to CLS. This gap can be described as the ignorance of the equivalence of two procedures. For aboutness, it is the ignorance of whether topic modeling detects themes in a way that is equivalent to the human practice of determining the respective themes based on reading. This ignorance of equivalence has two dimensions: firstly, the internal dimension of the operative structure of the procedure itself, and, secondly, the external dimension of the results (Hammond 2017). For the former, the claim of ignorance means that there is no evidence that a quantitative procedure performs the same operative steps as human minds do. With regard to the second dimension, there is the problem that we do not know whether the results are equivalent because the results have different forms. In other words, the ignorance consists of the problem that the output of both procedures are incommensurable. To bridge the methodological gap we shall propose a ternary model of operationalization and validation, which is visualized in Figure 1. This model is more comprehensive and, as we shall demonstrate, more powerful than the established binary conceptions of validation. In this way, our contribution fundamentally differs from the general criticism of topic modeling as it has been put forward in recent criticism (Shadrova 2021) that rejects topic modeling based on the claim that the concept of topic in topic modeling would have to be identical to the concept of topic in the domain where the procedure shall be used (in our case, the concept of aboutness in the domain of literary studies), and on the finding that there is no such conceptual identity, i.e. no co-extensionality (identity of references) and co-intensionality (identity of definitions) between both (i.e. the concepts of topic and aboutness). We maintain that Shadrova's requirement is far too strong. It is true but also obvious that the notion of topic in topic modeling is not co-extensional and co-intensional to the concept of aboutness in literary studies. We rather seek to develop a strategy of applying topic modeling in a hypothesis-based design that allows to investigate aboutness independently of the notion of topic. We maintain that the following model facilitates such kind of hypothesis-based analysis.

Figure 1 shows three units (qualitative concept, annotated texts, and quantitative procedure) with three binary relations between each two of these units. These three relations, one between the quantitative procedure and the intension (i.e. the definition) of the qualitative concept, another between the qualitative concept and the annotated texts, and the final relation between the results of the quantitative procedure and the extension (i.e. the scope of objects the concept refers to) of the annotated texts, mark the locations where different kinds of validation are required. So far, discussion on validation usually has limited itself to one of these three relations respectively.¹ We maintain that a full understanding of the impact of topic modeling as a technique of analyzing aboutness in the context of hypothesis-driven research (and not only in that of exploring corpora) necessitates that all three relations be modeled and validated. In the following sections, we shall demonstrate the general methodological requirements for ternary validation

1. Swafford (2015) focuses on the relation between the intension of the procedure and the qualitative concept, Piper (2015) on that between concept and annotations.

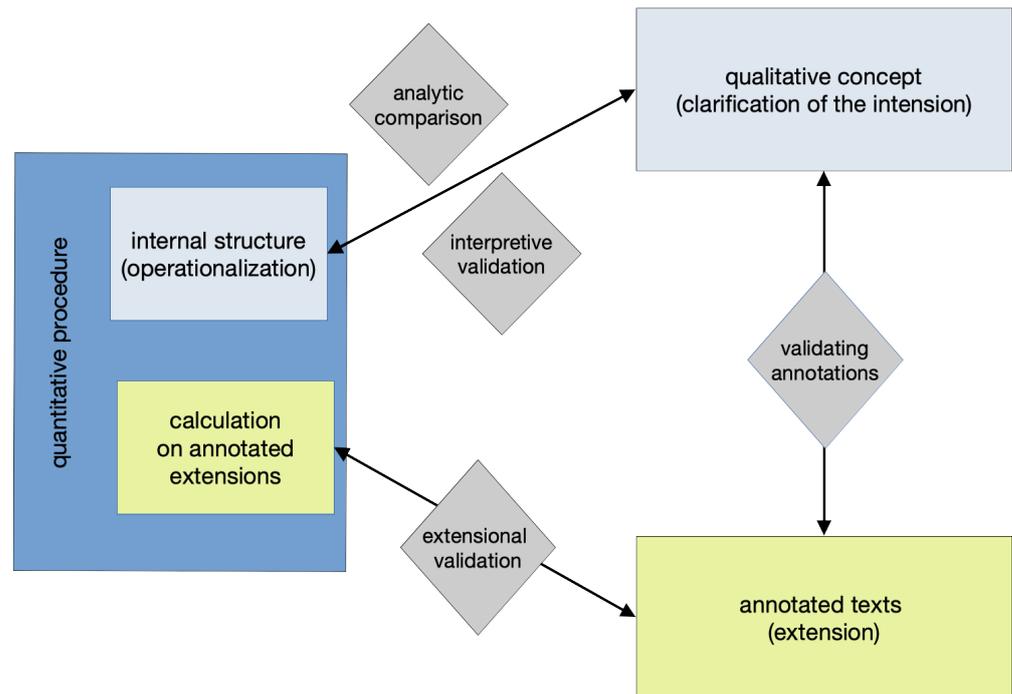


Figure 1: Ternary model of validation.

by discussing the three relations successively. Our methodological discussion will be empirically supported and illustrated by a new and large corpus of so far unknown 19th-century German novellas.²

2. Disambiguation and Internal Validation

We first discuss the relation between the intension of the qualitative concept and the quantitative procedure, which is, according to the first axiom, the relation between aboutness and the internal structure of topic modeling. The theoretical reason why we consider this relationship to be problematic is that we question the third axiom of the adequacy of distributional semantics. In other words, the following disambiguation shall demonstrate that topic modeling does not exactly reproduce aboutness in the way the concept of aboutness is used in literary studies. We do not contest that distributional semantics can be an appropriate and satisfactory theory within specific domains of linguistics, in particular for scenarios focusing on word similarity and synonymity or concerning usefulness in the context of information retrieval. From the perspective of literary studies, however, the distributional idea of semantics does not suffice to define the notion of aboutness because it can take different forms. We, therefore, have to think in scenarios of aboutness-claims. For this purpose, literary theory provides helpful terminological distinctions.

2.1 Conceptual Clarification: Aboutness as Sujet, Fabula, and Theme

The list of notions that are often used synonymously to indicate aboutness could be extended with 'subject', 'subject matter', or, in more specific contexts, 'issue', or 'problem'.

2. For a description of the corpus see the data repository.

Concerning the general grammatical structure, aboutness occurs as about-p-assertions such as ‘this novel is about love’. Two terminological distinctions from literary theory are relevant in the first instance, that between subject and theme (Lamarque 2009), and that between *sujet/syuzhet* and *fabula* in the tradition of Russian Formalism (introduced by Tomaševskij ([1931] 1985)), which has been translated to the distinction between story and plot in narratology. We take the latter distinction as a specification of Lamarque’s notion of subject so that we can focus on three terms: *fabula*, *sujet*, and theme. Tomaševskij defines *fabula* as the temporal and causal sequence of events. In large parts, this notion corresponds to that of Lamarque’s idea of subject: “To say what a work is about at subject level is in effect to retell the story or, in the case of non-narrative works, to redescribe the occasion or emotion presented” (Lamarque 2009, 150). In short, *fabula* is the plot-based aspect of aboutness. In contrast to *fabula*, both, Lamarque and Tomaševskij, define theme as the rather abstract unity of a literary work. This unity is, in most cases, not obvious but a result of interpretation. *Sujet*, which is a widely but heterogeneously used term in literary studies, is defined by Tomaševskij as the way the *fabula* is presented on the level of discourse including not only digressions, analepses and prolepses, as it has been emphasized in narratology, but also the setting (the place and situation of the *fabula*), the time, and the way characters are described (and, for example, dressed) and so on. In his illustrative analyses, Tomaševskij uses *sujet* to denote those aspects of the setting and surrounding that are not part of the *fabula* itself. In Aristotelian terms, *sujet* can in practice be used as the sum of the *accidentia* of the *fabula*.

We discuss the operationalizability of theme, *fabula*, and *sujet* based on the following illustrative extraction of several claims and interpretative hypotheses from different discursive contexts on one of the most canonical novellas of the period of Realism in 19th-century German literature, Keller’s *Romeo und Julia auf dem Dorfe* (1855/75):

- (a. love-1) The novella “treats the theme of love and death” (Saul 2003, 138).
- (b. love-2) The novella is about the tragic conflict between ideal, absolute, and unconditional love in contrast to social constraints (Kaiser 1971, 30).
- (c. love-3) The novella is about the problematic concept of love itself that has been internalized by the protagonists (Holub 1985, 476).
- (d. love-4) The novella is about structural incest in terms of Freud’s psychoanalytic theory (Holub 1985, 481).
- (e. *sujet*) The novella is an instance of the set of texts that are located in a rural surrounding (Stocker 2007, 72), it takes place “in an isolated rural ‘Dorfgeschichte’ location” (Saul 2003, 133).
- (f. social-1) The novella is about a devastating destiny caused by a violation of ownership (Menninghaus 1982 according to Walter Benjamin).
- (g. social-2) Based on the symbolic meaning of the character of the black fiddler, the message of the novella is that “in all members of the community [...] is an inner Gypsy, in all those secure in their unreflected homely identity lies hidden the exotic other” (Saul 2003, 139).

(h. structure) The aesthetic value of the novella results from reflexivity on semiotic processes and intertextuality, which is a step from realism to aestheticism (Stocker 2007, 69–75, Saul 2003).

The first claim (a) is an aggregation of fabula that is extended in the subsequent claims on the theme of love to a more complex structural thematic claim. All but (e) and (h) are aboutness claims. The latter does not point to the theme but the semiotic structure of the text. The contrast between the love claims (a to c), the psychoanalytic thesis (d), and the claims on social issues (f and g) shows that thematic claims are often controversial, sometimes absurd, and, in all cases, the result of intensive interpretive work. The claim on sujet (e) is a description of the text regarding general literary forms. As there is a tendency in literary studies towards giving interpretations of theme a higher prestige than analyzing sujet,³ we shall address the possible objection that claims on sujet are not aboutness claims in a proper sense. It seems to be clear that Keller's novella is a tragic love story but not a village story. This objection implies that aboutness relates to theme or fabula, but not to sujet. It is, however, also true that the novella is about love in a rural setting. Hence, sujet can be part of aboutness claims. Such claims have the logical structure 'x is about p in setting s'. As p refers to fabula or theme in claims of that type, theme, fabula, and subject can be nested. Our illustrative example at the end of this paper demonstrates that sujet can be significant to literary history, too.

2.2 Comparing Procedure and Conceptual Intension

The first relation that has to be validated requires the operationalization of a technical procedure that promises to approximate the conceptually clarified notion of aboutness. We distinguish three steps of operationalization, (1) that of selecting a promising quantitative technique or method, (2) that of adjusting factors that could impact the output of the selected procedure, which includes not only the parameters of the algorithm but also operations such as preprocessing textual data, and, if topic modeling is used, (3) that of selecting promising candidate topics. The subsequent fourth step is commonly labeled as 'internal validation' (Hammond 2017). It would be equally correct to label this type as 'intensional validation' because the internal structure of a quantitative procedure is compared to the intension of a qualitative concept from literary studies.

We start discussing internal validation concerning sujet:⁴ Several sujets such as surrounding, furnishing, or dressing, that are denoted by a limited set of descriptive terms or named entities, can be expected to be expressed satisfactorily by lists of weighted words. Romanesque environment, which is relevant to German novellas, can be expected to be approximated by words including named entities of cities or regions.⁵ Another relevant sujet, that of a 'rural surrounding', can be expected to be expressed by nouns that denote typical buildings or the specific social structure in villages, or nouns and verbs that express or refer to typical activities such as agriculture. Prior to validation, the degree of strength between a specific word and sujet should be taken into account in terms of a theory of meaning. Of course, the occurrence of words is neither

3. Lamarque (2009), who highlights the relevance of eternal and universal themes for assessing literary value, is representative of the tendency in literary studies to regard thematic interpretation as the more prestigious task compared to analyzing sujet and fabula.

4. Existing research occasionally interpreted concrete topics as indications of sujet (Schöch 2017), but did not yet provide a theoretical account of the relationship between topic modeling and sujet.

5. 'Romanesque environment' means that it is fictional that the story is located either in France, Italy, or Spain.

sufficient nor necessary for any sujet in a strict sense because lists of weighted words are not the proper representation of sujet but rather an approximation. Named entities, however, which are proper names in contrast to general terms,⁶ are almost inevitable for an author if a story shall be located in a certain setting. It is hardly possible to tell a story that takes place in Paris without referring to the name “Paris” or to entities that clearly refer to places, buildings, well-known events, or prominent historical persons in Paris.⁷ This strong relationship between named entities and sujet, which can be expected for a Romanesque setting does not hold for the sujet of a rural surrounding because it has to be approximated by general terms rather than named entities. Therefore, a heuristic distinction between sujets that shall be approximated mostly by singular terms and sujets that shall be approximated by general terms is useful for estimations prior to validation. Such estimation will also instruct the process of operationalization and of preprocessing because it requires that named entities are not removed from the corpus. According to Tomaševskij ([1931] 1985, 220), local or dynamic sujet, which is present only in particular scenes of a story, can be distinguished from global or static sujet that is prevalent over the whole text. The former requires that the texts be split up into segments. Prior to validation, we can assume that topic modeling performs best for stereotypical and homogenous global sujets that are approximated by named entities. The more local or dynamic and the more abstract and heterogeneous a sujet, the smaller the chances of success and the higher the efforts for parameter adjustment and for text manipulation in the process of preprocessing.

The third step is that of selecting the *prima facie* best topics after generating a topic model. This step is necessary because of two restrictions: Firstly, as the previous paragraph demonstrated, not all sujets can be expected to be approximated by topics. Secondly, not all topics are good candidates for approximating specific sujets⁸. Fortunately, topic modeling is capable of returning several promising village topics for our 19th-century novella corpus. The most promising candidate (topic no. 64, see code repository) starts with the nouns “Dorf” (village), “Haus” (home), “Mann” (man), “Knecht” (servant), “Leute” (people), “Feld” (field), “Wald” (forest), “Wagen” (carriage), “Pferd” (horse), “Bauer” (peasant), “Stall” (barn), “Arbeit” (labor). These words may create the impression of a good approximation to the sujet of a rural surrounding. For the sujet of a Romanesque surrounding, however, we were not able to identify any promising candidate topic. The occurrence of the names of cities, regions, or other entities that refer to French, Italian, or Spanish surroundings is not distributed with sufficient frequency and density in the text. In place of topic modeling, we developed another method of generating lists of semantically related words by manually drawing up a list of expected words such as “France”, “Italian”, or “Naples”, and determining the 50 nearest vectors to each of the words in the initial list, based on a SpaCy language model. Then, we summed

6. This distinction can be traced back to Frege (1892).

7. We can, of course, think of counterexamples. The story of Flaubert’s *Madame Bovary*, for example, is not located in Paris but the main character Emma often thinks of Paris and longs for living there. The novel has 75 hits for Yonville, the village where the action takes place, 74 hits for Rouen, the town that serves Emma as a replacement for her desire for Paris, and 34 hits for Paris. Of course, it would be mistaken to infer that the story is situated for 20% in Paris and, respectively 40% in Rouen and Yonville, as the absolute word counts suggest. It is nevertheless true that the novel is, in part, about a female protagonist’s thoughts about Paris.

8. This latter limitation, that a considerable number of topics in each topic model does not approximate semantic content but rather condenses rhetorical, stylistic expressions or verbs of communications, etc., is well reflected in all studies on topic modeling and expressed by the distinction between interpretable and non-interpretable topics.

up all nearest vectors for all words and selected the 30 most frequent words, which yield the final embedding-based list. Then, for all texts in the corpus, we calculated the relative share of this list by counting all lemmatized words of the text that are in the respective list and dividing by the sum of all word tokens in the text.⁹ Although both the village topic as well as the embedding-based list can be expected to be competing for approximations to specific *sujets*, no reliable insight is gained unless the techniques are validated also with regard to the remaining two relations.

With *fabula*, things are more complicated than with *sujet*. As *fabula* is defined as the causal progression of events, it implies a change in situation. In the case of love stories, events of falling in love are followed by a threat to the love relationship, and, finally, either by the elimination of the threat or of the failure of love. As for *sujet*, the proper representation of *fabula* is not a list of words, but rather a summary. Recently, more advanced methods of automatic summarization have been developed. “Automatic summarization seeks to present given information in a more compact form, determining the key messages of the text and eliminating unnecessary details and filler sentences” (Alexandr et al. 2021). The earlier approaches are mostly focused on extracting key sentences or passages as the summary of a document (Neto et al. 2002; Ribeiro et al. 2013). Such approaches have improved thanks to the recent development of deep-learning-based pre-trained language models. By identifying the key concepts and entities in the source document, automatic summarization combines the word-embedding-based representation of the input document and other linguistic features such as part-of-speech and named-entity tags (Nallapati et al. 2016). For its automatic evaluation, ROUGE (Recall-Oriented Understudy for Gisting Evaluation) has been suggested in Lin (2004). The idea is to count the overlapping textual units between the generated summary and a set of gold reference summaries. For the human evaluation of automatic summarization, Kryściński et al. (2019) suggested that the summaries should be evaluated from four perspectives: Coherence, Consistency, Fluency, and Relevance. Based on this instruction, automatically generated summaries are rated by human annotators on a Likert scale.

Such methods of summarization can be expected to be better approximations to *fabula* than topics. As this paper focuses on the scope of topic modeling, we can ask, nonetheless, whether several plot structures have semantic consistency over the whole text irrespective of situative changes during the progression of events. Although topics and other types of word lists are not proper representations of *fabula*, there can be pragmatic reasons for using word lists as rough approximations to static kinds of *fabula* such as crime, love, Western, or seafaring stories.¹⁰ As this holds only for several plot structures, the rationale for this consistency has to be reflected in terms of semantic theory: In many love stories, the aspect of love can be expected to be present globally over the whole text. For love stories, it is not the setting but rather the mode of communication and its characteristic forms of address that justify prior assumptions of semantic unity on the level of word lists. For stories about seafaring, western (Jannidis et al. 2019, 169), and several other highly stereotypical plots, in turn, it is not the plot structure itself that is represented in the topic, but rather the global *sujet* of a surrounding that is strongly connected to the plot. According to the terminological disambiguation we introduced

9. Code and the resulting lists are documented and explained in the code repository.

10. The latter is an example in Jockers (2013, 125). In our topic model, there is a highly conspicuous seafaring-topic (no. 98), too.

in this section, it would be more appropriate to say that there are several text types such as Western or Seafaring stories that are characterized by a specific plot structure as well as by a specific global sujet. In such cases, topic modeling does not identify fabula but rather sujet, which are, however, connected to fabula in the case of specific genres.

For theme, things are even more complicated than with fabula and sujet. Our illustration of interpretative claims on Keller's novella shows that several abstract concepts can serve as an abbreviation either for a typical plot structure or for thematic theses, where two operations can be observed: In our example, the core concept of love is integrated into the structural claim that there is a conflict between love and another abstract entity. Moreover, claims (b) and (c) indicate that one of several different general ideas of love is actualized in the text: in (b) that of radical and absolute romantic love, in (c), in contrast, that of not sufficiently radical love. The scope of both claims can only be understood properly if competing concepts of love are held present in the horizon of expectation. We refer to one of the most advanced theories of semantic change, Luhmann's *Liebe als Passion* (Luhmann 1982), which distinguishes (1) idealized love in medieval culture (*fin amour*), which is based on ratio based idolization mediating the difference between animalistic sexuality and sublime love, (2) paradoxical passionate love based on the idea of kurtosis and excess (*amour passion*), (3) love as friendship, (4) romantic and radically individualized love that is not concentrated on the character of the beloved, but on self-referential love itself, (5) the trivialization and ideology of reproduction where love as passion and romantic love appear as a problem that is transformed towards comradeship so that love becomes a matter of matrimonial viability mediating between the individual and social restraints.

Schöch (2017) in his study on the correlation between topic and genre identifies three different love topics that correlate with different dramatic sub-genres. When he notes that "each of the 'love' topics actually represents quite a different perspective on the theme of love", he interprets these candidate topics as representations of different abstract ideas of love, for example, "love as challenge and reward". Based on our terminological disambiguation, we can see more clearly that this exploratory strategy of interpreting topics starting from the resulting word lists is ambiguous. It may be the case that different love topics indeed approximate different abstract ideas of love. It is, however, also possible, according to the correlation between topic and genre that Schöch verifies, that all love topics refer to the same abstract idea of love but rather indicate different courses of fabula: One topic may include words that refer to a tragic ending whereas another refers to a happy ending. It is likewise possible that different love topics refer to different sujets such as different surroundings (for example, love in a rural versus urban milieu). Different topic word lists that are semantically related to love do not provide any information as to whether that topic approximates different concepts of love or different sujets or fabula aspects. For the general semantic relation between word lists and aboutness, we assume the following relation: the stronger the process of abstraction from fabula and sujet to theme and the more complex the propositional structure of thematic claims, the smaller the chances of success that thematic claims can be represented in lists of weighted words. Hence, we should not expect topic modeling to reveal thematic claims.

id	topic words	words that do not belong to vil- lage topic	interpretable as village topic
28	alt tag alte hoch gut kapitel bamme rot be- ginnen seidentopf groß hohen-vietz herz nehmen bild tür jung dorf schritt hohen- vietzer	alt, tag, alte, gut, kapitel, ...	No
64	dorf haus mann hof knecht leute schloß kom- men rufen feld sagen wald förster wagen ste- hen pferd bauer sehen stall arbeit	haus, schloß, rufen, sagen, ste- hen, sehen	Yes
38	dorf mühle hand ameile fränz schauen haus welt marann furchenbauer mund sagen gehen bauer frau vater hof stube munde bruder	ameile, fränz, schauen, ...	Yes

Table 1: Questionnaire of manual evaluation of topics.

2.3 Interpretive or Intensional Validation

If topic modeling shall be applied in the context of testing hypotheses regarding the presence of specific sujets or concepts at the core of themes, a further step of interpretive validation after operationalization is common practice (e.g. Navarro-Colorado 2018; Rhody 2014). We illustrate this strategy by adapting it to the case of rural surroundings in correlation with different candidate topics. According to current evaluation strategies (Aletras and Stevenson 2013; Mimno et al. 2011; Newman et al. 2010), topics can be manually evaluated through a questionnaire. Table 1 shows in an illustrative manner the first lines of such a questionnaire for three candidate topics of a rural surrounding. A common scenario for the application of interpretive validation in Digital Humanities is that people acquire a rough knowledge of several texts of an object area with a rough idea of typical, sujets, fabulae, and themes (in our example the knowledge of novellas and the idea that some novellas are about love, some are situated in a rural surrounding, etc.). Each row in the questionnaire contains the 20 most frequent words of the respective topic. One task is to identify all words that do not belong to the respective sujet, fabula, or theme. The other task is to decide whether the respective topic words approximate the annotator's qualitative notion of the respective sujet, fabula, or theme.

Two coefficients can be calculated from this type of questionnaire: Firstly, a ranking of words that are most often expected for village topics across all evaluated topics and all annotators, secondly, the average number of the minimum of words that must belong to a topic of a specific sujet, fabula, or theme can be determined. In this way, an empirical link between topics and qualitative concepts on the level of intension can be achieved. We have to concede here that such validation is much more complicated for more complicated sujets or concepts of love. For different ideas of love, expected words have to be articulated in advance. For *fin amour* in Luhmann's terms, descriptions of perfection, and expressions of admiration have to be expected in combinations with articulations of being in love. For *amour passion*, descriptions and expressions of passion as well as of feigning love are to be expected, for romantic love the singularity of the love itself, and for love as companionship nouns that express or denote friendship and descriptions of the reality of matrimonial and family live.

Irrespective of the practical difficulties for more complex sujets and themes, there are, however, to our mind, critical shortcomings of this strategy if it shall be transferred to the domain of literary studies: The presented type of evaluation has been developed within

and for computational linguistics according to its proper needs: “For our purposes, the usefulness of a topic can be thought of as whether one could imagine using the topic in a search interface to retrieve documents about a particular subject” (Newman et al. 2010). This particular strategy has then been adapted to the specific domain of information retrieval and relies on a rather restricted idea of the usefulness of topics. In the domain of information retrieval, this strategy may be appropriate. In the realm of literary studies, however, readers are more likely to adjust their expectations concerning aboutness to the presented lists of weighted topic words in a way that departs from the way they would estimate the presence or absence of specific sujets or themes if they were not confronted with topic word lists. Although the presented type of interpretive validation seems to be promising, it does not guarantee that the validated topics are actually about the respective sujet or theme, which is identified by close reading without looking at the results of quantitative procedures. Therefore, external validation is necessary.

3. Validating Annotations

The relation between the intension of a qualitative concept (such as *amour passion*) and the practice of identifying and annotating the presence of that concept in literary texts has to be clarified in an intermediate step. This clarification is not part of the quantitative procedures and of operationalization itself. In many scenarios, however, CLS cannot dispense with this dimension of validation (Schröter et al. 2021) and there is the possibility of validating this relation. There is, however, further need for a more systematic assessment of the methodologically controversial aspect of this type of validation. It is not entirely consensual how aboutness is represented in terms of reader response. Readers’ judgments with regard to the aboutness of literary works are, as Piper (2015) points out, subjective in general and often arbitrary or idiosyncratic. In such cases, there is, in statistical terms, high variance and low agreement between readers, which cannot be ignored as normal noise. As all people have different positions in the world,¹¹ Piper (2015) rightly stresses the *a priori* subjective character of readers’ judgments. If, however, judgments were completely arbitrary, reader response would be the expression of totally private feelings but not a response to texts as existing objects. From a pragmatic point of view, there are always fields of more consensual descriptions and there are domains of wider spread and lower inter-annotator agreement. Therefore, two further aspects have to be introduced. Firstly, the distinction between the psychological and the hermeneutic side of reader response. Secondly, the scaling from the intensional subjectivity of single annotations to extensional intersubjectivity.

For the first aspect, the dimensions of epistemic genesis and epistemic validity have to be distinguished. Concerning validity, aboutness is relevant either as a mental representation in concrete readers or as an objective property of a text as an entity. With regard to epistemic genesis, in contrast, aboutness is measured based either on empirical reader-response analysis or expert judgement or technical procedures. This dual distinction of validity and genesis is represented in Table 2, which records proponents and opponents of the possible positions.

11. This is what Davidson (2001, 39) calls the rational and unproblematic form of relativism in contrast to conceptual and epistemological relativism.

genesis – validity	empirical study	reader-response	hermeneutic reasoning	technical procedure
insight into the object itself ^a (objectivism)	Mellmann and Willand (2013) as proponents; rejected as ‘psychologism’ by Frege (2021) and Husserl ([1900] 2009).		Lamarque (2009)	Carnap (1950) (cf. Schröter et al. 2021)
insight into a perspective on objects (perspectivism)	Piper (2015)		relativist or constructionist professional reading, Barthes (1971)	Underwood (2019)

a. We do not distinguish between the currently dominating nominalist version and the outdated perspective based on a realism of universals (Stegmüller 1969, XXI).

Table 2: Modeling the difference between epistemic genesis and validity.

Both objectivism and perspectivism are legitimate frames for different research interests. However, objectivist interests necessitate reasonable and regulated annotations, whereas perspectival interest makes sense only based on perspectival data. Perspectival data can, for example, be extracted from contemporary reception documents such as reviews, articles, diaries, or letters for historical cultures, or from annotations, interviews, or surveys for present cultures.

Concerning the second aspect, that of transforming subjective and intensional reader response to extensional and intersubjective judgments, things are different for the relationship between objectivism and perspectivism. For both, it will be essential to calculate the spread of inter-annotator agreement in order to assess the degree of intersubjective consensus versus subjective arbitrariness. Under an objectivist interest, the spread of inter-annotator agreement is a strong benchmark of validity of annotations. Low agreement between annotators is problematic because it shows that the intension of the concept that shall be annotated has either not been sufficiently clarified prior to the task of annotating or that it is not clear in itself. Hence, a high spread should lead to revising the intension and the rules of annotation. If inter-agreement cannot be achieved, external validation will not be possible.

For perspectival modeling, in contrast, a low agreement between historical agents indicates that the concept was not well defined in contemporary culture. In the specific design of perspectival modeling (Underwood 2019), validating the historical perspective concerning intensions is not necessary. It is, in general, not necessary if the meaning of the historical perspectives does not need to be articulated in analytical terms of literary studies. Validation is necessary, in contrast, if a historical practice or a quantitative procedure or both shall be expressed in terms of literary studies. This is the case for interpreting topic modeling as an approximation to *sujet*, *fabula*, and *theme*.

For operationalizing *sujet*, *fabula*, and *theme* as properties of texts and not as historical perspectives based on topic modeling, an objectivist design is necessary. For *sujet* and *fabula*, a higher inter-annotator agreement can be expected than for *theme*, which highly depends on abstraction and imports of external theories (such as psychoanalytical theory in the thematic claim d or historical materialism in claim f of subsection 2.1). For abstract ideas such as different concepts of love within structurally complex thematic claims, a sufficiently high agreement between annotators will require extensive training on foundational theories. For the validation study that we present in the final section, the *sujets* of a rural surrounding and Romanesque environment as well as the idea

qualitative sujet, fabula, or theme	Annotated samples (size)	quantitative approximation	t-test, t-statistic	t-test, p-value	classifica- tion (LR), accuracy score
rural surrounding	'located in a village' (46) versus urban milieu (56)	topic no. 64	1.899	0.061	0.511
		topic no. 38	1.233	0.222	0.404
		topic no. 28	-0.556	0.580	0.399
		list of words, based on em- bedding	2.962	0.004	0.616
Romanesque setting	'located either in Spain, France, or Italy' (25) versus 'located elsewhere' (78)	list of words based on em- bedding	5.542	5.448e-7	0.786
romantic love	a story featuring romantic love (82) versus stories that are not love stories (36)	topic no. 36	-0.587	0.559	0.401
		topic no. 47	2.951	0.004	0.627
		topic no. 34	3.211	0.002	0.628

Table 3: Extensional evaluation of rural surrounding, Romanesque setting, and romantic love.

of romantic love were disambiguated, in case of the latter concept according to Luhmann (1982) (see subsection 2.2), and transferred into rules for annotating about 100 novellas.¹²

4. External or Extensional Validation

The final and most important relation that has to be validated is that between the extension of the qualitative concept and the extension of the quantitative procedure. Hence, we shall refer to this type, which is sometimes called external validation in linguistics (Gries 2008, 427), as extensional validation. Based on annotations (or, in case of perspectival modeling, on reader-response analysis of reception evidence) as described in the preceding section, the extension of texts with a specific sujet, fabula, or theme in qualitative terms has to be provided and compared to the results of the quantitative procedure. There is an important restriction to this type of validation. As Shadrova (2021) points out, the results of this type of validation cannot be generalized. This is certainly true with regard to the inductive structure of empirical inference in general. In our case, the results for extensional validation of the quantitative procedure for operationalizing a specific sujet, for example Romanesque setting, cannot be generalized for the relationship between topic modeling and all sujets. Shadrova, however, over-emphasizes this restriction. We maintain that it is possible to articulate systematic hypotheses on generalizability based on specific empirical validations. Such hypotheses have to be proved in subsequent case studies. Hence, we shall present an example for an extensional validation and discuss possible generalizations in the conclusion of this paper. Our case study is based on our novella corpus and. Its results are recorded in Table 3. The disambiguated qualitative concept of the respective sujet or theme is recorded in the first column, its translation into samples based on annotations, according to the process of transforming intensions to extensions as elaborated in section 3, is recorded in the second column.

A methodological issue arises as we have to relate a categorical variable (presence or absence of a sujet, fabula, or theme) with a metric value of quantitative procedures.

12. The results are stored in the data folder in the code repository.

Accordingly, there are two options: The weaker and easier option is to calculate the share of the respective word list for the contrary groups based on annotation. According to the distribution (mean and standard deviation) of the dominance of words of the list in both contrary samples, a t-test (here Welch's t-test for samples with different variance) is calculated. Its t-statistic and p-value are recorded in the third and fourth column for scaled data. This first option is applicable in contexts of weak comparative hypotheses. The stronger the difference for the share between the contrary samples, the higher the probability that a high value for individual texts indicates that a text has the respective sujet, theme, or fabula recorded in the first column. The second option is more demanding and it is required in contexts, where the quantitative results, which are in their very structure metric, can be interpreted categorically in a way that a threshold facilitates classifying texts as having a specific sujet, fabula, or theme. For our examples, we performed a classification task with the metric value of the topic share or the word list share as the independent predictor variable and the qualitative sujet, fabula, or theme as the dependent predicted variable based on a logistic regression algorithm, with cross-validation and a custom-made bootstrapping method with 10,000 iterations of resampling, training, and calculating the accuracy scores for predications on a validation set. For each sample of contrary subsamples of the same size, with the larger subsample reduced to the size of the smaller subsample randomly, 80% of the documents were used for training and the remaining 20% for validation. The final column records the accuracy scores for predictions on the validation set. For comparison, we conducted a simple bag-of-words based classification to set a baseline. The classification for annotated sets of rural surrounding, Romanesque environment, romantic love are 0.401, 0.400 and 0.540 for the 5,000 most frequent and tf-idf normalized words as features, respectively.¹³

For the statistical significance of the hypothesis that both samples are from different populations (which means that texts with a specific sujet, fabula, or theme are different from texts without that sujet, fabula, or theme) as well as for the results of the classification task, we see that the candidate topics selected from our topic model performs better than the baseline of classifying annotated samples based on a document term matrix of the 5,000 most frequent tf-idf normalized word types but worse than our generated word lists based on word embedding. In a future study, we shall address the methodological ground for such embedding-based lists. With regard to our theoretical discussion in [section 2](#), we can understand why the Romanesque setting based on a list generated by word embedding has the best performance and why no candidate topic word for this sujet could be generated. Words that indicate Romanesque surroundings do not appear with sufficient frequency and equal dispersion in the texts concerning topic modeling. If such words (for example, named entities of cities and regions) appear in a text, however, these words are highly specific to and indicative of a Romanesque setting. Also for romantic love, the embedding-based word lists outperform topic modeling. For rural surroundings, the best candidate topic has the same performance as the embedding-based word list. If two sufficiently large annotated validation samples were available, a more refined strategy would be advisable. The first sample could be used as a test set in a grid search for optimizing parameters such as the total amount of topics, length of chunks, and hyperparameters of the algorithm itself. According to the results of the grid search, candidate topics with the best performance in the discussed

13. All details of the significance test and the classification are documented in the code repository.

classification task can be identified. With the second set as a validation sample, the optimized topic model could be validated as discussed in this section.

Against this proposed strategy of extensional validation, one could object that the aboutness of texts does not have to correlate with high dominance of specific topics. With regard to *sujet*, this objection can be appropriate because it can be necessary for local *sujets* to calculate the share of topic dominance not for whole documents but only for specific segments. In general, however, this objection amounts – intentionally or unintentionally – to the claim that topic modeling would be completely irrelevant concerning aboutness. If this objection holds true, the dominance of specific topics for singular documents would not have any meaning. It was the aim of this paper, however, to provide the ground for strategies that allow proving whether there is such a meaning of the dominance of topics with regard to the question of what texts are about.

5. Conclusion

This paper has a methodological impact as well as an empirical result: With regard to the first, we claim that it is common practice in CLS to distinguish between thematically interpretable and uninterpretable topics. This dichotomy of interpretability versus non-interpretability has two weaknesses: Firstly, it is imprecise because our disambiguation demonstrated that ‘theme’ (from Jockers 2013) often means ‘fabula’ or ‘sujet’ and that both notions refer to different types of textual properties. The second weakness is that it has not yet been validated whether topics really approximate specific *sujets*, *fabulae*, or concepts within thematic claims. In this paper, we maintain that validation is not, as methodological discussion in CLS suggests so far, either internal or external. It is rather located on a relation between (a) the intension and (b) the extension of a qualitative concept and (c) a quantitative procedure. On each relation of this triangle, conceptual clarification, explication, and operationalization are important methodological units and are interlinked with different tasks of validation. Hence, we do not claim that everything is validation or that validation is everything, but, rather, that validation pops up at all three relations of a holistic research design. Disambiguating different forms of aboutness is necessary and limiting oneself to specific aspects (such as certain *sujets*) is useful because quantitative procedures are expected to behave unequally to different *sujets*, *fabulae*, and themes so that different forms of aboutness need different operationalization and individual validation.

Although singular validation results cannot be generalized in a simple way and without further empirical proof, our illustrative example in the fourth chapter can serve as a starting point for generalizations that have to be proved in forthcoming studies. Based on rational reflection and the results of our case study, we expect *sujet* to be better operationalizable with topic modeling than *fabula*, and *fabula* to be operationalized in specific cases such as seafaring or Western as *sujet*. Such cases may be well operationalized because of their homogeneous setting, which is linked with *fabula* according to genre rules. In such cases, it is rather *sujet* than *fabula* that is represented by word lists. With regard to theme, only the isolated abstract concepts that have a basis on the level of *sujet* or *fabula* in a text (such as love) can be expected to be operationalized with word lists. We suggest that the practice of operationalization should be regarded as a

recursive process that repeatedly compares the intension of the qualitative concept with the internal structure of the quantitative procedure and adjusts the parameters of that procedure based on such comparison. Therefore, one of our potential future works is to test LDA with different parameter settings and also to test more advanced quantitative methods such as Deep Neural Networks-based topic models (Zhao et al. 2021) or state-of-the-art language models, to find out whether more complex aboutness-claims in literary corpora could be operationalized.

In technical terms, topic modeling reduces the dimensions of a document-term matrix of a corpus. Internal validation with reference to the intension of the qualitative concept is the most common and often an appropriate form of validation in computational linguistics. However, as we discussed in the paper, only some of the topics can be used as the representation of a small part of the distribution of aboutness in literary corpora. In CLS, internal validation can be useful but it is not sufficient because it does not guarantee that topics are capable of identifying texts that have the respective *sujet*, *fabula*, or *theme* from the perspective of hermeneutics. Our results for the extensional validation support this suspicion.

The empirical result is that, based on extensional validation, topic modeling did not perform with statistical significance in all cases. However, the calculated t-statistic has a positive value for all but one candidate topics, which implies that topics mostly indicate the expected tendency. From this empirical result, we can draw several hypothetical generalizations. We assume that topic modeling is not able to identify aboutness for all sorts of *sujet*, *fabula*, and *theme* in a strict sense. A two-step validation strategy based on two different annotated validation samples and a grid search for optimizing parameters could, however, yield better results for topic modeling in future research. As the discussion of conceptual intension and interpretive validation in [section 2](#) demonstrated, it is hardly possible to generate promising topics as approximations to *sujets* such as Romanesque setting. For other *sujets* that have promising approximations as topics, the method performs more poorly than the method generating lists based on word embedding. As simple word lists with equally weighted words are less complex than topics with differently weighted words, this result may be astonishing. Based on analytic reasoning and for the *sujet* of a Romanesque surrounding, however, this result comes as no surprise. Whereas existing studies examined the applicability of topic modeling in different domains (e.g. Navarro-Colorado 2018), we applied it to the domain of narrative fiction and come to the preliminary conclusion, that in the realm of analyzing aboutness topic modeling may be most appropriate to operationalize *fabula* related *sujets* such as Western or Seafaring because of the homogeneity of setting-references and the high frequency of these references. Non-*fabula* based *sujets* such as location in a specific cultural environment may be operationalizable with dictionary or word-embedding based word lists. These results do not reduce the applicability of topic modeling for domains different from aboutness, for example for analyzing historical (Lee 2019) or philosophical (Nichols et al. 2018) discourses. Therefore, we do not share Shadrova's general scepticism against the non-generalizability of topic modeling. If the statistical characteristics of each quantitative procedure are taken into account and related to the terminological definitions of philological notions of *fabula*, *theme*, and *sujet*, there is new epistemic ground for articulating hypothetical generalizations of the particular empirical results of validation studies. If these hypothetical generalizations

can be proved in further studies, stronger empirical evidence for the appropriateness of specific quantitative procedures for analyzing general types of aboutness can be gained.

6. Data Availability

Data can be found here: <https://github.com/julianschroeter/19CproseCorpus>

7. Software Availability

Software can be found here: https://github.com/julianschroeter/evaluating_topic_modeling_for_sujet_and_theme

8. Acknowledgements

The construction of the corpus and the task of annotation was performed by Theresa Valta, Johannes Leitgeb, and Julian Schröter and has been funded by the Forschungsfonds der Philosophischen Fakultät der Universität Würzburg from 2018 to 2020 in the course of a larger project on the history of German novellas (<https://www.germanistik.uni-wuerzburg.de/en/lehrstuehle/computerphilologie/mitarbeiter/schroeter/forschung-publikationen/habilitationsprojekt/>).

Parts of this contribution resulted from the project “A Mixed Methods Design for Computational Genre Stylistics and Unstructured Genres. Towards a Functional History of 19th Century German Novellas” (project number 449668519), which was funded by the *Deutsche Forschungsgemeinschaft* (DFG) as a Walter-Benjamin Fellowship.

9. Author Contributions

Julian Schröter: Conceptualization, Formal Analysis, Writing – Original Draft, Data Curation, Validation, Software

Keli Du: Methodology, Writing – Review & Editing, Formal Analysis, Validation, Software

References

- Aletras, Nikolaos and Mark Stevenson (2013). “Evaluating Topic Coherence Using Distributional Semantics”. In: *Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013) – Long Papers*. Association for Computational Linguistics, 13–22. <https://www.aclweb.org/anthology/W13-0102> (visited on 05/05/2020).
- Alexandr, Nikolich, Osliaikova Irina, Kudinova Tatyana, Kappusheva Inessa, and Puchkova Arina (2021). “Fine-Tuning GPT-3 for Russian Text Summarization”. In: *Proceedings of the Computational Methods in Systems and Software*. Springer, 748–757. [10.48550/arXiv.2108.03502](https://doi.org/10.48550/arXiv.2108.03502).
- Barthes, Roland (1971). *S/Z*. Seuil.

- Blei, David M. (2012). "Probabilistic Topic Models". In: *Communications of the ACM* 4 (55), 77–84. [10.1145/2133806.2133826](#).
- Blei, David M. and John D. Lafferty (2009). "Topic Models". In: *Text Mining: Classification, Clustering and Applications*. Ed. by Mehran Sahami Ashok Srivastava. Chapman and Hall/CRC, 71–94.
- Carnap, Rudolf (1950). *Logical Foundations of Probability*. University of Chicago Press.
- Da, Nan Z. (2019). "The Computational Case against Computational Literary Studies". In: *Critical Inquiry* 3 (45), 601–639. [10.1086/702594](#).
- Davidson, Donald (2001). "The Myth of the Subjective". In: *Subjective, Intersubjective, Objective*. Clarendon Press, 39–52.
- Evert, Stefan (2005). *The Statistics of Word Co-Occurrences: Word Pairs and Collocations*. Institut für maschinelle Sprachverarbeitung, Universität Stuttgart.
- Firth, John R. (1957). *Papers in Linguistics. 1934-1951*. Longmans.
- Frege, Gottlob (1892). "Über Sinn und Bedeutung". In: *Zeitschrift für Philosophie und philosophische Kritik* (100), 25–50.
- (2021). *Die Grundlagen der Arithmetik*. Wilhelm Köbner.
- Gries, Stefan Th. (2008). "Dispersions and Adjusted Frequencies in Corpora". In: *International Journal of Corpus Linguistics* 4 (13), 403–437. [10.1075/ijcl.13.4.02gri](#).
- Hammond, Adam (2017). "The Double Bind of Validation: Distant Reading and the Digital Humanities' 'Trough of Disillusionment'". In: *Literature Compass* 8 (14), 1–13. [10.1111/lic3.12402](#).
- Harris, Zellig S. (1954). "Distributional Structure". In: *Word* 2-3 (10), 146–162.
- Holub, Robert C. (1985). "Realism, Repetition, Repression: The Nature of Desire in *Romeo und Julia auf dem Dorfe*". In: *MIn* 3 (100), 461–497.
- Husserl, Edmund ([1900] 2009). *Logische Untersuchungen, Erster Band: Prolegomena zur reinen Logik (1900/1913)*. Meiner.
- Jannidis, Fotis, Leonard Konle, and Peter Leinen (2019). "Makroanalytische Untersuchung von Hefromanen." In: *DHd Book of Abstracts*, 167–173. [10.5281/zenodo.4622093](#).
- Jockers, Matthew L. (2013). *Macroanalysis: Digital Methods and Literary History*. University of Illinois Press.
- Jockers, Matthew L. and David Mimno (2013). "Significant Themes in 19th-century Literature". In: *Poetics* 6 (41), 750–769. [10.1016/j.poetic.2013.08.005](#).
- Kaiser, Gerhard (1971). "Sündenfall, Paradies und himmlisches Jerusalem in Kellers *Romeo und Julia auf dem Dorfe*". In: *Euphorion* 1971 (65), 21–48.
- Kryściński, Wojciech, Nitish Shirish Keskar, Bryan McCann, Caiming Xiong, and Richard Socher (2019). "Neural Text Summarization: A Critical Evaluation". In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 540–551. [10.18653/v1/D19-1051](#).
- Lamarque, Peter (2009). *The Philosophy of Literature*. John Wiley & Sons.
- Lee, Changsoo (2019). "How are 'immigrant workers' represented in Korean news reporting?—A text mining approach to critical discourse analysis". In: *Digital Scholarship in the Humanities* 1 (34), 82–99. [10.1093/llic/fqy017](#).
- Lin, Chin-Yew (2004). "Rouge: A Package for Automatic Evaluation of Summaries". In: *Text Summarization Branches Out*. Association for Computational Linguistics, 74–81.
- Luhmann, Niklas (1982). *Liebe als Passion: zur Codierung von Intimität*. Suhrkamp.

- Mellmann, Katja and Marcus Willand (2013). "Historische Rezeptionsanalyse. Zur Empirisierung von Textbedeutungen". In: *Empirie in der Literaturwissenschaft*. Ed. by Philip Ajouri, Katja Mellmann, and Christoph Rauen. Brill | mentis, 263–281. [10.30965/9783957439710_015](https://doi.org/10.30965/9783957439710_015).
- Menninghaus, Winfried (1982). "Romeo und Julia auf dem Dorfe. Eine Interpretation im Anschluss an Walter Benjamin". In: *Artistische Schrift. Studien zur Kompositionskunst Gottfried Kellers*. Suhrkamp, 91–158.
- Mimno, David, Hanna M. Wallach, Edmund Talley, Miriam Leenders, and Andrew McCallum (2011). "Optimizing Semantic Coherence in Topic Models". In: *Proceedings of the conference on empirical methods in natural language processing*. Association for Computational Linguistics, 262–272.
- Nallapati, Ramesh, Bowen Zhou, Cicero dos Santos, Çağlar Gülçehre, and Bing Xiang (2016). "Abstractive Text Summarization using Sequence-to-sequence RNNs and Beyond". In: *Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning*, 280–290. [10.18653/v1/K16-1028](https://doi.org/10.18653/v1/K16-1028).
- Navarro-Colorado, Borja (2018). "On Poetic Topic Modeling: Extracting Themes and Motifs From a Corpus of Spanish Poetry". In: *Frontiers in Digital Humanities* (5). [10.3389/fdigh.2018.00015](https://doi.org/10.3389/fdigh.2018.00015).
- Neto, Joel Larocca, Alex A. Freitas, and Celso AA Kaestner (2002). "Automatic Text Summarization Using a Machine Learning Approach". In: *Brazilian symposium on artificial intelligence*. Springer, 205–215.
- Newman, David, Jey Han Lau, Karl Grieser, and Timothy Baldwin (2010). "Automatic Evaluation of Topic Coherence". In: *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, 100–108.
- Nichols, Ryan, Edward Slingerland, Kristoffer Nielbo, Uffe Bergeton, Carson Logan, and Scott Kleinman (2018). "Modeling the Contested Relationship between Analects, Mencius, and Xunzi: Preliminary Evidence from a Machine-Learning Approach". In: *The Journal of Asian Studies* 1 (77), 19–57. [10.1017/S0021911817000973](https://doi.org/10.1017/S0021911817000973).
- Piper, Andrew (2015). *Validation and Subjective Computing* [Blog post]. <http://txtlab.org/?p=470> (visited on 04/22/2022).
- Rhody, Lisa (2014). "The Story of Stopwords: Topic Modeling an Ekphrastic Tradition." In: *Book of Abstracts DH2014*.
- Ribeiro, Ricardo, Luís Marujo, David Martins de Matos, Joao P. Neto, Anatole Gershman, and Jaime Carbonell (2013). "Self Reinforcement for Important Passage Retrieval". In: *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*. Association for Computing Machinery, 845–848.
- Saul, Nicholas (2003). "Keller, *Romeo und Julia auf dem Dorfe*". In: *Landmarks in German Short Prose*. Ed. by Peter Hutchinson. Peter Lang, 125–140.
- Schöch, Christof (2017). "Topic Modeling Genre: An Exploration of French Classical and Enlightenment Drama." In: *Digital Humanities Quarterly* 2 (11). <http://www.digitalhumanities.org/dhq/vol/11/2/000291/000291.html> (visited on 11/08/2022).
- Schröter, Julian, Keli Du, Julia Dudar, Cora Rok, and Christof Schöch (2021). "From Keyness to Distinctiveness – Triangulation and Evaluation in Computational Literary Studies". In: *Journal of Literary Theory* 1-2 (15), 81–108. [10.1515/jlt-2021-2011](https://doi.org/10.1515/jlt-2021-2011).

- Shadrova, Anna (2021). "Topic models do not model topics: epistemological remarks and steps towards best practices". In: *Journal of Data Mining & Digital Humanities* (2021). [10.46298/jdmdh.7595](https://doi.org/10.46298/jdmdh.7595).
- Stegmüller, Wolfgang (1969). *Wissenschaftliche Erklärung und Begründung*. Springer-Verlag.
- Stocker, Peter (2007). "Romeo und Julia auf dem Dorfe. Novellistische Erzählkunst des Poetischen Realismus". In: *Gottfried Keller. Romane und Erzählungen*. Ed. by Walter Morgenthaler. Reclam, 57–77.
- Swafford, Annie (2015). *Why Syuzhet Doesn't Work and How we Know*. [Blog post]. <https://annieswafford.wordpress.com/2015/03/30/why-syuzhet-doesnt-work-and-how-we-know/> (visited on 04/22/2022).
- Tomaševskij, Boris ([1931] 1985). *Theorie der Literatur. Poetik*. Seemann.
- Underwood, Ted (2019). *Distant Horizons: Digital Evidence and Literary Change*. University of Chicago Press.
- Weitin, Thomas and Katharina Herget (2017). "Falkentopics. Über einige Probleme beim Topic Modeling literarischer Texte". In: *Zeitschrift für Literaturwissenschaft und Linguistik* 1 (47), 29–48.
- Zhao, He, Dinh Phung, Viet Huynh, Yuan Jin, Lan Du, and Wray Buntine (2021). "Topic Modelling Meets Deep Neural Networks: A Survey". In: *arXiv preprint*. [10.48550/arXiv.2103.00498](https://arxiv.org/abs/10.48550/arXiv.2103.00498).