

Topic Modeling for the Identification of Gender-specific Discourse

Virtues and Vices in French and Spanish 18th Century Periodicals

Yvonne Völkl¹ 
Sanja Sarić² 
Martina Scholger² 

1. Institute of Interactive Systems and Data Science (ISDS), Graz University of Technology , Graz, Austria.
2. Institute Centre for Information Modelling – Austrian Centre for Digital Humanities, University of Graz , Graz, Austria.

Citation

Yvonne Völkl, Sanja Sarić, and Martina Scholger (2022). "Topic Modeling for the Identification of Gender-specific Discourse. Virtues and Vices in French and Spanish 18th Century Periodicals". In: *Journal of Computational Literary Studies* 1 (1). [10.48694/jcls.108](https://doi.org/10.48694/jcls.108)

Date published 2022-11-24

Date accepted 2022-03-30

Date received 2021-12-22

Keywords

topic modeling, French, Spanish, Spectator Press, 18th century, Literary Gender Studies

License

CC BY 4.0 

Reviewers

José Eduardo González 

Glenn Roe 

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. Gender-specific knowledge – just like knowledge in general – is generated through discourses that are disseminated through (mass) media. Among the first mass media is the Spectator press (*Moralische Wochenschriften*) which spread all over Europe throughout the 18th century. With their gender-specific discourses, analyzed in *Spectatoriale Geschlechterkonstruktionen* (Völkl 2022), they decisively promote the development of a (bourgeois) gender model, shaping the social perception of gender until today. Against this background, the present article examines the gender-specific discourses in the French and Spanish Spectator periodicals by means of topic modeling which detects semantically related words. The study, which originates from the project *Distant Spectators. Distant Reading for Periodicals of the Enlightenment* (Scholger et al. 2019–2021), shows that topic modeling reinforces previous findings on gender-specific discourses in the Spectator periodicals. Moreover, it offers new perspectives concerning this research corpus.

1. Introduction

The Spectator periodicals are a popular journalistic genre of the 18th century which (co-)constructs and preserves the cultural knowledge of its time in general and gender-specific knowledge in particular, propagating a heteronormative society. As a broadly effective medium circulating from England throughout the Western world, the Spectator periodicals also promote the transcultural dissemination of a transforming understanding of gender,¹ in conjunction with the changing values, norms, and practices among the constituting middle classes.

1. In the course of the 18th century, the perception of female and male bodies and their genitalia changed. The so far gradually assumed difference between women and men is increasingly interpreted qualitatively and a complementary understanding of two genders can assert itself (cf. Laqueur 2003[1990]). The new perception of women and men also leads to a cultural redefinition of their gender relations (e.g., woman as the 'moral gender', cf. Steinbrügge 1987) and to a major shift in the conception of virtue which was originally connotated to the meritorious properties and qualities of men (Latin: *vir*) and was feminized as of the end of the 17th century only (cf. Pabst 2007, 25ff.).

The quantitative-statistical as well as the discourse-analytical and interpretative study on gender-specific ways of worldmaking in the French- and Spanish-language Spectator periodicals (cf. Völkl 2022) reveals that the French-language periodicals of the first half of the 18th century contribute to the dissemination of the notion of a ‘natural’ gender difference which primarily appears together with a discourse of character and/or physical differences. From the middle of the century onwards, in which the periodicals are also published in Spain,² the discourse of difference is expanded to include the aspect of complementarity, finally recognizing woman and man as a mutually complementary entity. Due to her alleged closeness to nature, in this discourse of complementarity, the woman is hierarchically placed under the authority of man, whose assumed higher ability to reason is considered superior.

In order to disseminate the discourses of difference and complementarity, the French- and Spanish-language periodicals draw particularly on the notion of virtue (French: *vertu*; Spanish: *virtud*). According to research on the Enlightenment period, this term generally functions as a gender-specific key concept (‘geschlechtsspezifischer Leitbegriff’ according to Pabst 2007) and stands in opposition to the notion of vice (French: *vice*; Spanish: *vicio*) (cf. Bolufer Peruga 1998, Kilian 2002, Schaufler 2002, Steinbrügge 1987). Furthermore, the discourse on virtues and vices is combined with positive and negative (character) traits and behavioral patterns which are hierarchized and assessed as worthy or not worthy of emulation. Among the ignoble vices on the one hand, one can find, e.g., hypocrisy, idleness, vanity, or jealousy which should be avoided by women and men alike (and thus remain gender-*un*specific). The virtues worthy of emulation on the other hand, are constructed in a gender-specific way, with the ‘female’ virtues revolving around concepts such as decency, modesty, kindness, shamefulness, beauty, or (a specific female) education, while the ‘male’ virtues only include (a specific male) education, honesty, and reason. In order to make the large number of virtues and vices known to the Spectator audience – which decidedly also included women – they are incorporated into gender-stereotypical models, illustrating ideal images or warning examples. E.g., the characteristics of egoism and vanity which are considered vicious, are linked to the stereotypical models of the coquette or the fop and contrasted with virtuous models of women and men. The gender-stereotypical models with their manifold virtues and vices are enveloped into countless (exemplary) stories from everyday life and in (character) portraits which are narratively woven into the plot (cf. Völkl 2022, 282–286).

To quantitatively verify these observations on the (entire) Spectator corpus, a topic modeling analysis was carried out in the course of the project *Distant Spectators. Distant Reading for Periodicals of the Enlightenment (DiSpecs)* (Scholger et al. 2019–2021) after which special attention was given to the interpretation of those topics that stand out from a gender-theoretical perspective. The computed values and their visual representation were intended to provide a new perspective on the corpus and create new theories and questions. The following chapters first describe the related work, the research corpus, and the methodology before presenting the results and findings of the topic modeling analysis with regard to gender-specific discourse in the Spectator periodicals.

2. For a description of the Spanish Spectator periodicals and an in-depth analysis of the use and function of the letter as mode of communication with the public, see Hobisch 2017.

2. Related Work

Topic modeling has become an integral part of the range of methods used in digital humanities, and more specifically in computational literary studies. According to the survey of Du, it has been increasingly used since 2011 (cf. Du 2019). In the field of historical newspapers and periodicals, topic modeling was conducted for analyzing the social and political life of Civil War Richmond based on the *Richmond Daily Dispatch* (cf. Nelson 2020) and for investigating the discourse dynamics in historical newspapers published in Finland between 1854 and 1917 (cf. Marjanen et al. 2020). Regarding the Enlightenment period, Schöch applied topic modeling on French Classical and Enlightenment drama for sub-genre classification (cf. Schöch 2017), and Roe et al. analyzed the discursive structure in the *Encyclopédie* of Denis Diderot and Jean le Rond d’Alembert (cf. Roe et al. 2016).

A persistent point of criticism in the application of topic modeling is the lack of explainability and comprehensibility of the results (cf. Hu et al. 2014, 424–425, Liu et al. 2017, 1–2). This is very much related to the lack of documentation of single working steps and parameters applied in the topic modeling process as Du pointed out (cf. Du 2019): In order to guarantee the reproducibility of the results, it is crucial to have details on the number of documents, the conducted pre-processing steps, the number of topics and iterations selected in the actual modeling process, etc. To address this criticism, this contribution aims to not only provide the results of our topic modeling analyses, but also to transparently document the workflow that led to them.

3. The Research Corpus

While the Spectators have been studied through close reading as a work-centered approach, there have been no previous activities that explore this genre from a distant reading perspective. For this reason, the project DiSpecs engaged in text mining of the collection of 3,863 periodical issues in six languages,³ assembled and edited during the digital scholarly edition project *The ‘Spectators’ in the International Context* (Ertler et al. 2011–2021). In the DiSpecs project, topic modeling was used for investigating the semantic and stylistic structure.

What proved to be very useful for the analysis was the fact that the documents were already available in XML/TEI format (TEI Consortium 2021). This includes not only the annotation of metadata and structural elements (e.g., paragraphs and pagination), but also narrative forms (e.g., self-portrait, letter/letter to the editor, fable) and narrative levels of representation,⁴ as well as subjects (e.g., ‘Idea of man’, ‘Nature’, ‘Economy’, ‘Theatre Literature Arts’), mentioned places, person names, and intellectual works. The

3. The corpus contains periodicals in English, French, German, Italian, Portuguese, and Spanish, but due to the rather small corpus size of English, German, and Portuguese, these languages were not considered in the topic modeling analysis. Therefore, we analyzed 1,658 French-, 1,344 Italian-, and 690 Spanish-language issues.

4. The Spectator periodicals stand out for their multi-layered system of communication consisting of various narrative levels of representation which are embedded in various narrative forms. Further, narrative forms are also intertwined within each other when, e.g., the fictitious editor includes a supposedly authentic reader’s letter in the periodical which, in turn, narrates a story about a woman who then enters into a dialogue with another woman about a letter to a man (cf. Fischer 2014, 81–83). This epistolary correspondence between editor and readers has been considered a major element for the success of the Spectator periodicals in the course of the 18th century (cf. Hobisch 2018).

annotation format provided through the application of the Text Encoding Initiative (TEI) standard enables easier extraction of certain structures of the data for the analysis (e.g., metadata, headings, footnotes, editorial comments) with the possibility to separate issues into paragraphs and to exclude parts or whole issues during the pre-processing of the data which will be explained closer in subsection 4.2.

4. Topic Modeling Workflow

The unsupervised probabilistic topic modeling method aims to identify hidden thematic structures in large text collections (cf. Blei 2020, 8) which means that the algorithm recognizes patterns in the data without having a training subset or a desired output (cf. Alloghani et al. 2020, 4). The resulting topics usually consist of thematically related words, i.e., tokens. However, some topics have structural rather than thematic significance. They can provide insight into the writing style of the author, terms typical for a genre, adjectives describing a matter, repeatedly mentioned places or persons. This is due to the fact that the method's algorithms measure the co-occurrence of words, following statistical assumptions, meaning that if the same words often occur together in documents, they are most likely thematically related (cf. Blei 2020, 9).

Multiple algorithms were developed for topic modeling, but one of the most prominent, and the one we used in our analysis, is Latent Dirichlet Allocation (LDA). We owe this choice to the DARIAH-DE team who developed a Jupyter Notebook embedding the *dariah_topics* Python library for topic modeling (DARIAH-DE 2019) with MALLET (McCallum 2002–2018), a toolkit that builds on LDA. We adapted and expanded these notebooks to incorporate them into our topic modeling analysis workflow,⁵ which can be divided into four main parts: data evaluation, pre-processing of the data, topic modeling creation and post-processing of the results (Figure 1). As we demonstrate in this chart, individual steps of the workflow have to be repeated to optimize the results. Further on in this chapter, we will describe how we conducted these steps and what decisions were important for quality results.

4.1 Data Evaluation

To get an overview of the French- and Spanish-language research data, we conducted a number of exploratory data analysis steps. This included evaluating and visualizing the size of the corpus, the number of issues per periodical and per author, the number of tokens per issue as well as the distribution of manually assigned keywords and narrative forms. This simple statistical analysis allows insight into the corpus which can be relevant for interpreting and evaluating the results. For example, comparing manually assigned keywords with topics identified through topic modeling is used for cross-evaluation of these two approaches, i.e., finding out how similar the human- and the machine-assigned topics are. In addition, discrepancies in the metadata could be detected and corrected. Getting this insight was possible thanks to the TEI annotation of the Spectator corpus which allows extracting all the relevant data structures from the documents either with Python libraries like *Beautiful Soup* or with XSLT while

5. The Jupyter Notebooks with the Python code are provided via GitHub: <https://github.com/distantpectators/DiSpecs>.

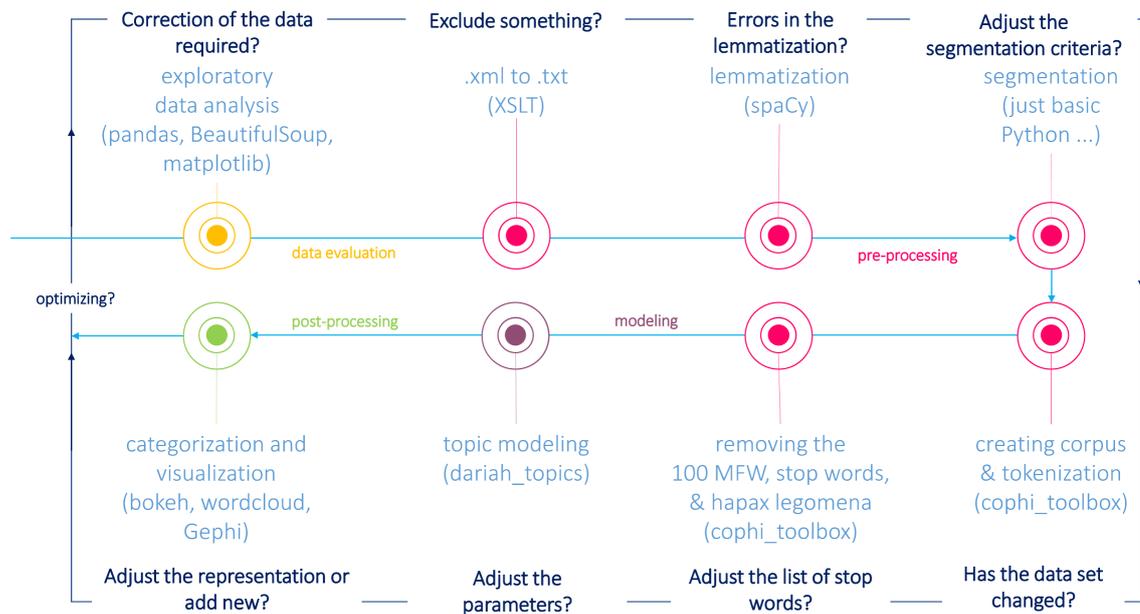


Figure 1: The topic modeling workflow.

transforming the XML/TEI to plain text files.

4.2 Pre-processing

Our workflow, building on DARIAH-DE, required plain text files as input for topic modeling. As part of this transformation, we extracted metadata from the TEI files and used it to build file names: publication year, periodical name, author of the periodical, volume, issue, and persistent identifier of the file. This way, we had easy access to specific parts of the files' metadata even when using plain text files. We also filtered the text material. On the one hand, we divided the collection into separate corpora according to their language and excluded files that did not contain manually assigned keywords (e.g., tables of contents). On the other hand, we removed titles and subtitles since they were repetitive and therefore had a disproportionate impact on the result.

These plain text files were already fulfilling formal requirements to proceed with topic modeling, meaning they were classified per language, in the desired format, and containing metadata in the file names. But LDA treats inflected forms (e.g., Span. *mujer* – 'woman', and *mujeres* – 'women', or *muger/mugeres* in 18th century orthography) as different concepts. A topic can therefore include multiple forms of the same concept which can result in semantically poor topics. To avoid this, we decided to lemmatize the Spectator texts before modeling the topics, using natural language processing with *spaCy* to replace each inflected word (e.g., *mugeres*) with its lexical base form (*mujer*), i.e., lemma. This step was one of the most challenging since *spaCy* was not trained on historical language. Wrongly lemmatized tokens had to be replaced with the correct

lemma through a dictionary. Although still not without errors, the decision to lemmatize brought much cleaner and semantically richer results than the preliminary experiments with non-lemmatized texts.

Since topic modeling measures the co-occurrence frequency of tokens in the same document, another pre-processing step was to define what will be treated as a document. We decided to segment the issues in paragraphs with a minimum of 500 tokens, whereby longer paragraphs were avoided by cutting off a paragraph after the first following sentence's end, if the number of included tokens had surpassed 600. Remaining paragraphs with less than 200 tokens were appended to the preceding paragraphs of the same issue, to avoid very short paragraphs. Although there were still a few outlier paragraphs left, this method resulted in a larger quantity of documents with a similar token number instead of a smaller quantity of documents with more strongly varying token numbers. Since there is no state-of-the-art consensus on the optimal number of tokens in a document, the selection was based on preliminary experiments with different values.

With this set of resized and lemmatized documents, we continued with the workflow as provided by DARIAH-DE, with some practical adjustments. From the imported and tokenized documents, we removed redundant tokens as a last pre-processing step since some tokens do not have semantic significance or are simply irrelevant and therefore not desired to be part of the final result. These tokens are a) the 100 most frequent words (MFW) because they tend to be functional words, like pronouns, articles, prepositions, etc., b) the hapax legomena (tokens occurring only once in the corpus), and c) a project specific stop word list. To create the stop word lists, we adjusted the *Stopwords ISO* (Diaz 2016) lists and expanded them after each of our topic modeling cycles with new resulting topic keywords we identified as irrelevant.⁶

4.3 Topic Model Creation

Since topic modeling is an unsupervised machine learning method, the researcher cannot impact the result by assigning categories in advance. There are, however, a couple of factors that do influence the results. One of them is, as previously mentioned, the pre-processing of the data. Another one is the choice of the input parameters: the number of topics, the number of iterations and the hyperparameter optimization interval. Table 1 gives an overview of relevant parameters in our topic modeling analysis.

The number of topics is thus decided by the researcher. The reasonable number of topics in a text collection depends on the text scope, but also the genre and the thematic richness. Our approach was to experiment with different numbers of topics and evaluate the results to decide the optimal number of topics for each text corpus. Eventually, we determined 25 French and 18 Spanish topics. The number of topic keywords, on the other hand, is not a matter of the researcher's decision: each topic consists of all tokens from the text collection, whereas the distribution of these tokens varies in the individual topics. So, each token from the treated text collection can be found in each resulting topic, but with a varying probability which is never equal to 0% (cf. S. Bock et al. 2016,

6. Besides functional words, such as the Spanish *inmediatamente* (immediately) or *entonces* (therefore), or some frequently used adjectives and modal verbs like the French *grand* (great) and *devoit* (should), we also excluded some nouns from the analysis, e.g., the Spanish *número* as it is often used as an issue title.

	French	Spanish
Number of periodicals	25	18
Number of issues	1,658	690
Extracted segments	6,752	3,190
Lemmatization	yes	yes
Removed features	100 MFW, 801 stop words, hapax legomena	100 MFW, 823 stop words, hapax legomena
Number of topics	25	18
Iterations	2,000	2,000
Alfa hyperparameter	5.0 (MALLET default)	5.0 (MALLET default)
Beta hyperparameter	0.01 (MALLET default)	0.01 (MALLET default)
Hyperparameter optimization	20	20

Table 1: Parameters used in the topic modeling analysis of French- and Spanish-language periodicals.

13). The researcher familiar with the analyzed content then decides on how many of the topic tokens, i.e., keywords, they see as significant to represent in the results. For each analyzed group, we chose to output the first 20 tokens.

The researcher also sets the number of iterations. More iterations lead to a longer processing time but can lead to more reliable and stable results until a limit is reached after which the quality stagnates (cf. Jockers 2014, 147). Choosing an optimization interval is optional and depends on the desire to observe the difference in topic weight, by “allowing some topics to be more prominent than others” (McCallum 2002–2018).⁷ In our analysis, we conducted 2,000 iterations with an optimization in every 20 iterations.

But even with the same data and the same parameters, the output of two modeling cycles is never exactly the same in terms of the topics per document and the words per topic distribution, due to the probabilistic and unsupervised nature of the method. Nevertheless, using our data and parameters, the comparison of multiple results showed a re-emergence of the same topics, with rather insubstantial differences in the sequence of the most frequent topic keywords as well as the probability of the topics which suggests a sufficient stability of the model.

4.4 Post-processing

The last step in the topic modeling workflow is the post-processing of the results. The probability of topics is being computed for each individual document (which, as explained in subsection 4.2, is a segment of a periodical’s issue). Using these values, we computed the probability of topics per periodical and represented the results from different perspectives, utilizing multiple visualization techniques.⁸

A common way to represent topics are heat maps. The heat map (Figure 2) is a visual representation of the data frame matrix (Table 2) resulting from the topic modeling and the computed results per periodical, consisting of periodicals (X axis), topics (Y

7. Schöch gave a more detailed reflection on hyperparameter optimization in his scientific blog (cf. Schöch 2016).

8. The heat map, bar charts, and word clouds of the French periodicals can be viewed under the URL <https://gams.uni-graz.at/o:dispeccs.result.tm.fr>. Spanish results are accessible under the URL <https://gams.uni-graz.at/o:dispeccs.result.tm.es>.

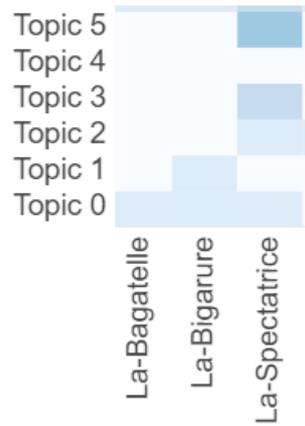


Figure 2: Heat map detail of 3 periodicals.

	La Bagatelle	La Bigarure	La Spectatrice
⋮	⋮	⋮	⋮
Topic 5	0.0196085	0.01663	0.126381
Topic 4	0.011347	0.0313603	0.03801
Topic 3	0.004227	0.00814	0.08031
Topic 2	0.03328	0.0153783	0.05184
Topic 1	0.02778	0.0434557	0.01843
Topic 0	0.064768	0.0685915	0.04833

Table 2: Data frame matrix detail corresponding to the heat map detail in Figure 2.

axis), and the probabilities of each topic per periodical as values where darker color represents higher probability.

For each periodical as well as for each topic, we created a bar chart (e.g., Figure 3). This technique offers a focus on one periodical or topic, whereas the heat map is better suited for getting an overview of the whole corpus. Both the heat map and the bar chart creation are part of the original DARIAH-DE Jupyter Notebook.

Additionally, we decided to use word clouds to visualize the top 100 keywords of the respective topics. A larger font size indicates a keyword with a higher probability inside a topic (Figure 4). We chose this visualization method despite some critics claiming that it is difficult for users to infer the relationship of the words from it (cf. Dobson 2021, §20). While we do agree with this point, a visual overview of all topic keywords is beneficial next to a visualization of the topic distribution, especially when the topics are not labeled.

Another way we used Python libraries to represent topics in selected periodicals is with line diagrams which show the prevalence of topics over the issues of a single periodical, i.e., over time. This is only a relative prevalence over time since the time span between the issues was not always constant and is not explicitly available in the metadata. The interactive diagram (created using the library *bokeh*) can be viewed on

our project website.⁹ Here, it is possible to zoom in, create sections, activate or deactivate the visibility of individual topics, and save the created versions of the diagram.

Finally, we used the software *Gephi* to create networks of topics, periodicals, and manually assigned keywords (Gephi.org 2008–2021). More precisely, we created a force-directed graph using the algorithms Fruchterman Reingold and Force Atlas 2 (Figure 7).¹⁰ The periodical nodes are represented as pie charts showing the distribution of a certain manually assigned keyword throughout the periodical's issues. The web presentations include color legends and numerical data for the pie charts. The size of a periodical node (pie chart) indicates whether the number of analyzed periodical issues from the topic modeling set is larger or smaller in comparison to other periodicals. Note that numerous issues do not mean the same as a large amount of text since some issues can be very long while others are quite short. The size of the topic nodes indicates whether a topic has a high or low representativity in the analyzed set of texts. The edges are higher weighted (thicker) if the likelihood of a topic in a periodical is higher. Nodes with the same color belong to the same community. This means that the densities of the edges between these nodes are higher than from these nodes towards the rest of the network. But, since this is a small network where all topics occur in all periodicals to some degree, the weighted modularity of this network is low, and the community structure is not perfectly clear. Nevertheless, it is possible to detect topics that often co-occurred in periodicals.

As shown by the visualizations, the topics are non-semantically labeled (Topic 0, Topic 1, Topic 2, ...), and the numbers give no statement about the importance or frequency of the topic but are only used to distinguish the topics. This approach is contrary to the occasionally seen practice where researchers either label their topics by interpreting them (cf. Boyd-Graber et al. 2017, 40, or Blevins 2010) or by using a few of the most relevant keywords, as proposed by the DARIAH-DE Jupyter Notebook. In recent years, we noticed an increase in the non-semantic labeling approach (cf. Horstmann and Kleymann 2019, Krautter et al. 2020, or Chehal et al. 2021). We also decided to proceed without labels because the interpretation of a topic depends on the reception horizon of the researcher. This further impedes the obtrusion of a certain perspective and leaves room for different interpretations. We did, however, provide our interpretation in textual form. The gender-specific topics will be elaborated in section 5 and section 6.

To ensure transparency and comprehensibility of the visualizations and interpretations, all the underlying raw data can be downloaded by the user, including the topic keywords list and the word weights. Nevertheless, it has to be pointed out that for understanding the results of distant reading, a certain familiarity with the source material through close reading expertise is always required to create meaning from the results and generate added value for related research. As Shadrova also suggests, “[i]t is of crucial importance to make the underlying contextualization, the model, explicit, both through hypothesis-based work and by tying results back to the theoretical and conceptual debates in the field” (Shadrova 2021, 16).

9. Line diagrams: <https://gams.uni-graz.at/archive/objects/o:dispecs.result.tm.fr/methods/sdef:TEI/get?mode=diachronic> and <https://gams.uni-graz.at/archive/objects/o:dispecs.result.tm.es/methods/sdef:TEI/get?mode=diachronic>.

10. The full visualizations can be viewed on our web page: <https://gams.uni-graz.at/archive/objects/o:dispecs.result.tm.fr/methods/sdef:TEI/get?mode=topic-network> and <https://gams.uni-graz.at/archive/objects/o:dispecs.result.tm.es/methods/sdef:TEI/get?mode=topic-network>.

Topic 4	Topic 9	Topic 18	Topic 21	Topic 22	Topic 24
fille	vertu	heureux	aimer	dame	air
jeune	mérite	dieu	sentir	quoiqu	bel
pere	vie	doux	bonheur	égard	sexe
mariage	nature	oeil	passion	manière	dame
fils	ame	tendre	lettre	passion	beauté
mere	hommes	tendre	amant	tem	jeune
mari	propre	main	tendre	sexe	visage
famille	bonheur	ciel	moment	mauvais	oeil
père	vice	ame	malheureux	propre	plaire
âge	passion	voix	douleur	convenir	aimable
enfan	heureux	feu	sentiment	peine	mode
marier	conduite	aimable	perdre	conduite	habit
chevalier	action	charme	malheur	obliger	joli
épouser	sage	gloire	heureux	penser	figure
tendresse	honneur	peine	tendresse	montrer	femmes
demoiselle	digne	objet	devenir	dessein	goût
devenir	noble	bel	ame	affection	conversation
maison	mal	terrebeauté	oeil	devenir	grace
soin	fortune	sage	objet	liberté	rire
amant	estime	brillant	étois	avis	compagnie

Table 3: Gender-specific topics in French-language Spectators.

5. Topic Modeling in the French-language Spectator Periodicals

Among the 25 topics of the Spectator periodicals published in French language, at least six topics stand out from a gender-specific perspective. Topics 4, 22, and 24 directly, topics 9, 18, and, 21 indirectly relate to character, behavior, and roles of women and men within the (emerging bourgeois) society in the 18th century (see Table 3).

Topic 4 lists the various French terms for ‘marriage’ and ‘getting married’ (*mariage*, *marier*, *épouser*), ‘family’ (*famille*), ‘child’ (*enfan*), or ‘house’ (*maison*) which are terms that construct the destiny of young (!) women (*fille*, *demoiselle*) within the domestic sphere (in contrast to the public sphere which is attributed to men).¹¹ In this private sphere, her main duty is to take tender (*tendresse*) care (*soin*) of her husband (*mari*) and children.

Topic 22 refers to the vocabulary used in the translation of the *Female Spectator* (1749–51), *La Spectatrice, traduite de l’anglais* (1750–51) as indicated in the bar chart with a probability of over 0.3 within this periodical (Figure 3) which is much higher in relation to other periodicals. *La Spectatrice* is one of the few spectatorial titles specifically directed to (bourgeois) women. This focus on the female readership also reverberates in the first term of the topic with ‘lady’ (*dame*). The subsequent terms used, such as ‘passion’ (*passion*), ‘bad’ (*mauvais*), ‘suitable’ (*propre*), ‘corresponding’ (*convenir*), ‘conduct’ (*con-*

11. Regarding gender discourse in 18th century France, see the articles of G. Bock and Zimmermann 1997, Brink 2008, Honegger 2011, or Sieuzac 2009. As to the presence of women in society and literature, see the essay collection edited by Jacobs et al. 1979. Concerning the theoretical and literary discourse on the woman as the ‘moral gender’ in the 18th century, see Steinbrügge’s monograph (Steinbrügge 1987). As to the representation of women in the French Enlightenment press, see Dijk 1988 and to the history of the ‘presse féminine’ in France, see Sullerot 1966.

duite), or ‘affection’ (*affection*), indicate that this topic is concerned with the behavior of women in public, especially in the company of men.

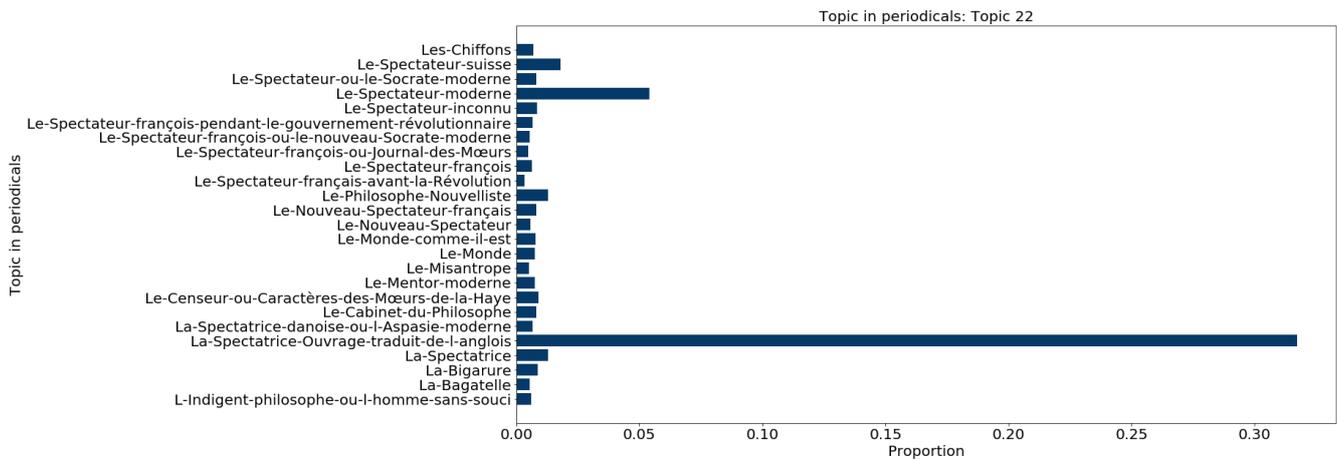


Figure 3: Distribution of topic 22 in French-language periodicals.

Topic 24, visualized as word cloud (Figure 4), also lists attributes associated with the ‘fair sex’ (*beau sexe*).¹² On the one hand, a woman has to ‘please’ (*plaire*) through her inner beauty – expressed by terms such as ‘beautiful’ (*bel*), ‘beauty’ (*beauté*), ‘amiable’ (*aimable*), and ‘grace’ (*grâce*) – and on the other hand through her outer beauty – expressed as well by the terms ‘beautiful’ (*bel*) and ‘beauty’ (*beauté*), but also by ‘pretty’ (*joli*) or ‘taste’ (*goût*). Both inner and outer beauty are accentuated by appropriate ‘clothing’ (*habit, mode*), good ‘taste’ (*goût*), and ‘conversation practices’ (*conversation*) that are understood as suitable for a woman. Her ‘appearance’ (*air*), i.e., her outward appearance, has the highest priority here, as can be seen from the prominent position of the term in the first place, and is represented in all periodicals (see also topic 17 of the Spanish periodicals where the orientation on outward appearances manifests through terms such as *moda* – ‘fashion’, *adornar* – ‘to adorn’, *gustar* – ‘to please’, *hermoso* – ‘beautiful’, *hermosura* – ‘beauty’).

The discourse on women within the French-language periodicals is further supported by topics 9, 18, and 21. While the first three topics mentioned above explicitly evoke terms for women (e.g., *beau sexe, femme*), and also use self-explaining terms alluding to their status (e.g., *dame* – ‘lady’, *demoiselle* – ‘unmarried young woman’, *mère* – ‘mother’) as well as gender-specific, heteronormative practices (e.g., *marier, épouser* – the act of getting married), the terms used in topics 9, 18, and 21 are more implicit to the extent that they only indirectly allude to the gender-specific discourse and roles of women and men in the (bourgeois) society within the Spectator periodicals.

The terms occurring in **topic 9** describe virtuous behavior and practices. The gender-specific key concept of virtue (cf. Pabst 2007) stands at the very beginning of the word sequence. The following terms refer to the fact that virtue leads to (individual and collective) ‘happiness’ (*bonheur*). In general, 18th century philosophers equate ‘virtue’ with ‘happiness’, for only those who lead a virtuous life can contribute to their own

12. The French term *beau sexe* for the female part of the population is a compound and has been separated during the topic modeling process. Further, the term *beau* has been lemmatized into *bel*. This is the reason why the terms *bel* and *sexe* appear separately in topic 24. Nonetheless, their immediate position next to each other indicates their connection.

Topic 24 in French Spectators

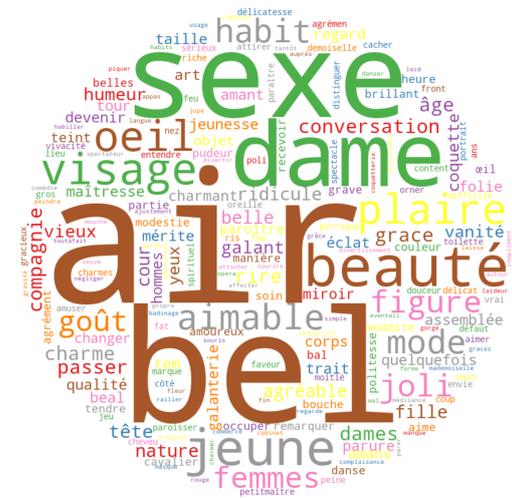


Figure 4: 100 MFW in topic 24 in French-language periodicals.

happiness and to the happiness of the community. Virtue is thus seen as a means to achieve the individual and collective goal of happiness (cf. Völkl 2022, 121–122).¹³

Moreover, as explained at the outset of this article, the concept of virtue becomes a gender-specific key concept in the course of the 18th century insofar as not only different virtues emerge for women and men, but also considering that women have to follow a larger amount of virtuous (i.e., morally acceptable) behavioral norms than men. On the discursive level, this appears through a quantitatively higher number of terms attributed to ‘female’ behavior than those attributed to ‘male’ behavior. The reason for this is that in the 18th century, the virtuousness of the whole (bourgeois) society becomes more and more associated with the virtuous behavior of women. In the Spectator press, precisely this virtue is constantly promoted (cf. Völkl 2022, 120–124).

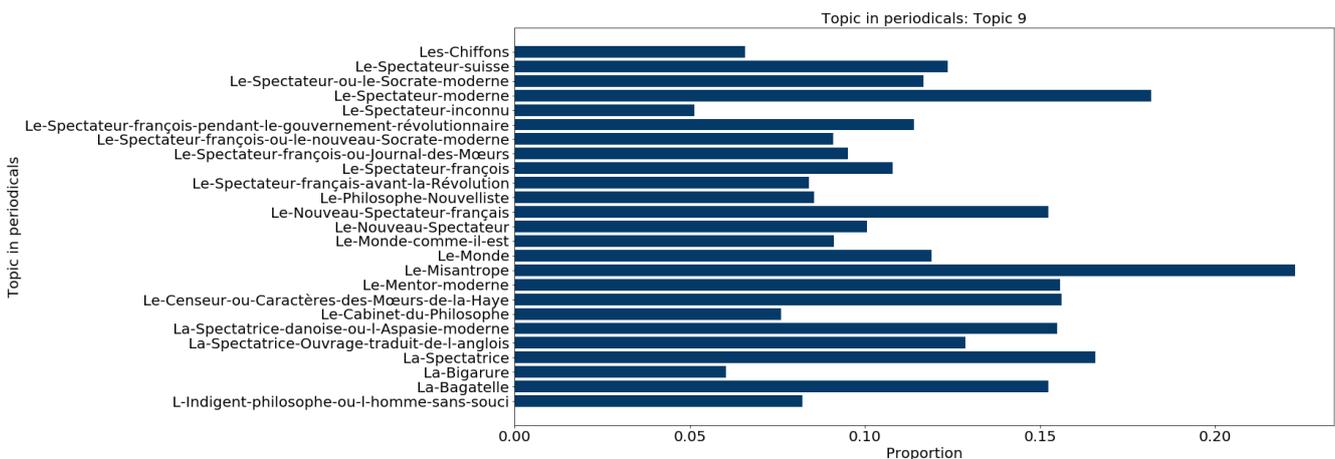


Figure 5: Distribution of topic 9 in French-language periodicals.

Topic 9 can be found in all Spectator periodicals at a median rate of 0.62 (Figure 5)

13. On the discourse of happiness in the 18th century, cf. Mauzi 1969, on the concept of ‘happiness’ in *The Spectator*, cf. Norton 2015.

which makes it the most probable in the corpus. This is not surprising because most periodicals explicitly state their goal in their first lines which is to turn all people into useful members of the society – a society which is becoming increasingly complex and integrated into a nation (cf. Ertler 2010, 100). In terms of women as useful members of society, the role of the (bourgeois) woman is conceptualized in three ways: as spouse, housewife, and mother. Outside the domestic sphere, she has no right to exist, which is the reason why, for example, the image of the learned woman was defamed in the Spectator periodicals at the beginning of the 18th century and has subsequently been omitted altogether – according to the motto ‘out of sight, out of mind’ (cf. Völkl 2022, 309–310).

Topic 18 results in terms referring to the virtuous ideal image of both women and men. The terms ‘tender’ (*doux, tendre*), ‘amiable’ (*aimable*), ‘grace’ (*charme*), ‘prudent’ (*sage*), ‘witty’ (*brillant*) here refer to inner virtues while the terms ‘beautiful’ (*bel*) and ‘beauty’ (*beauté*) can refer to inner and outer virtues at the same time as explained above. Although this is not a frequent topic, it is consistently present in all Spectator periodicals.

Topic 21 exhibits terms that can be assigned to the discourse field of love. They are associated positively or negatively with love. For example, next to the approbatives ‘to love’ (*aimer*), ‘happiness’ (*bonheur*), or ‘tender’ (*tendre*), one can find the pejoratives such as ‘unhappy’ (*malheureux*), ‘pain’ (*douleur*), or ‘to lose’ (*perdre*). The frequency of individual terms will be discussed below.

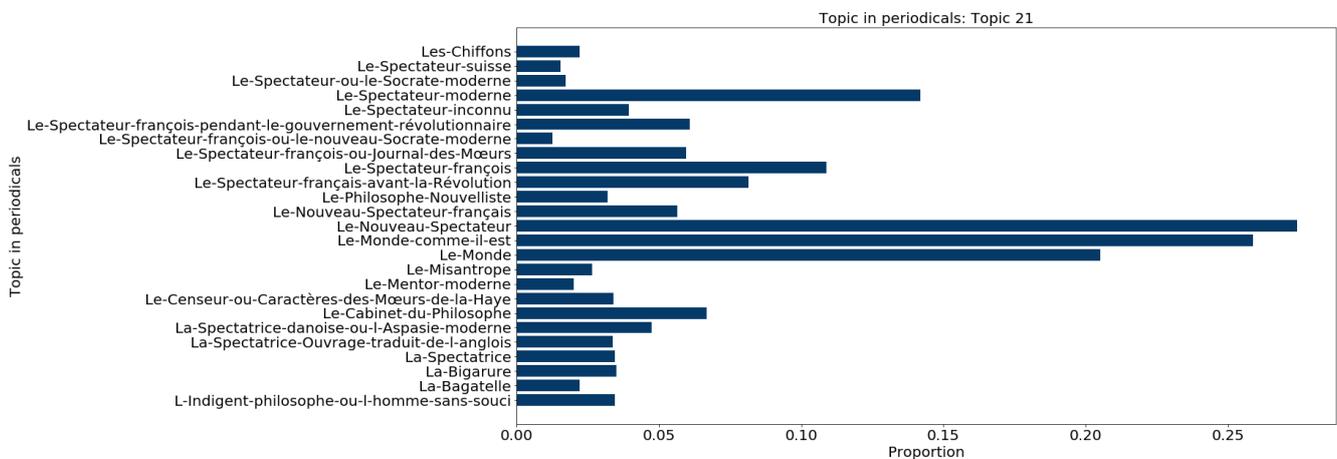


Figure 6: Distribution of topic 21 in French-language periodicals.

A look at the distribution of topic 21 within the French-language periodicals (Figure 6) reveals that the three successive periodicals *Le Nouveau Spectateur* (1758–60), *Le Monde comme il est* (1760), and *Le Monde* (1760–61) of Jean-François de Bastide (1724–1788) are particularly endowed with this topic. The literary and cultural studies research carried out by Fischer-Pernkopf et al. and Völkl support the finding that Bastide continuously narrates exemplary stories of happy and unhappy (heterosexual) lovers (cf. Fischer-Pernkopf et al. 2018, Völkl 2022).

The proximity of the nodes and the edges weight in the network analysis graph in Figure 7 also illustrates the prevalence of topic 21 in all of Bastide’s periodicals. It further shows

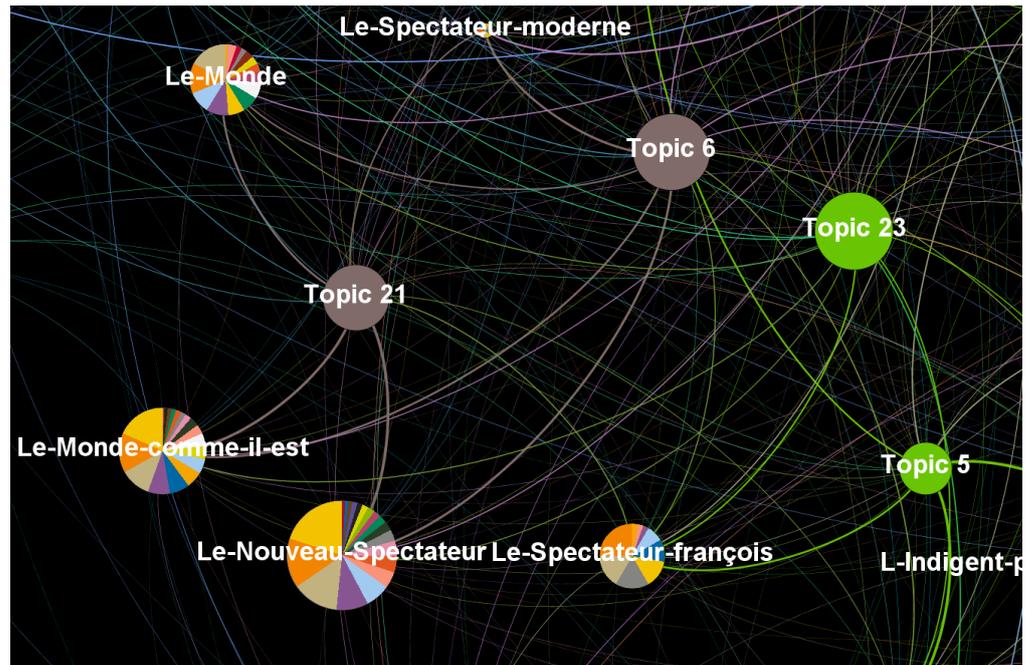


Figure 7: Detail from the network graph of periodicals and topics, showing the prevalence of topic 21 in the three periodicals of Jean-François de Bastide.

that **topics 6**¹⁴ and **23**¹⁵ are also very common in Bastide's periodicals. They include typical narrative vocabulary (e.g., *demander* – 'to ask', *répondre* – 'to answer', *entrer* – 'to enter', *entendre* – 'to listen', *lire* – 'to read') and typical narrative elements (e.g., *ami* – 'friend', *maison* – 'house', *chambre* – 'room', *porte* – 'door'). Based on the accumulation of narrative terms, it can be concluded that in Bastide's periodicals, the discourse of love is primarily conveyed through stories and storytelling. This interpretation of the topic modeling results is supported by previous literary and cultural research in this field which also stresses the strong narrative design of Bastide's periodicals (cf. Fischer-Pernkopf et al. 2018, Mussner 2016, Völkl 2022).

Additionally, the analysis of the issues manually annotated with the subjects/keywords 'Image of women' and 'Image of men'¹⁶ of the *Nouveau Spectateur*¹⁷ and the *Monde comme il est*¹⁸ identified that the following five narrative forms (*Erzählformen*) are predominately used to discuss family life (in particular education) and couple relationships (with a focus on the romantic tender love relationship): general account (*allgemeine Erzählung*, AE), heteroportrait (*Fremdporträt*, FP), metatextuality (*metatextueller Kommentar*, MT), dialogue (*Dialog*, D) and letter/letter to the editor (*Leser*innenbriefe*, LB) (cf. Völkl 2022, 209). Concerning the distribution and arrangement of these text types, it has to be emphasized that they also repeatedly appear intertwined within each other which leads

14. Topic 6: penser vérité mal vrai honneur ami réflexion juger mauvais défaut répondre caractere sentir droit convenir lorsqu connoître entendre lire quelquefois

15. Topic 23: maison demander heure chambre paraître jeune tem main peine passer entrer revenir air ami vouloit porte arriver alloit entendre sortir

16. In total, the list of subjects comprises 37 keywords which were determined at the beginning of the digital scholarly edition project (cf. Ertler et al. 2011–2021) and which was slightly expanded in the course of the project.

17. From the 108 issues within the *Nouveau Spectateur* 58 issues (44%) are indexed by the subject 'Image of women' and 16 issues (14,8%) by the subject 'Image of men' (cf. Völkl 2022, 206).

18. Within the 60 issues of Bastide's *Monde comme il est*, 38 issues (64%) are indexed by the subject 'Image of women' and 12 issues (20%) by the subject 'Image of men' (cf. Völkl 2022, 206).

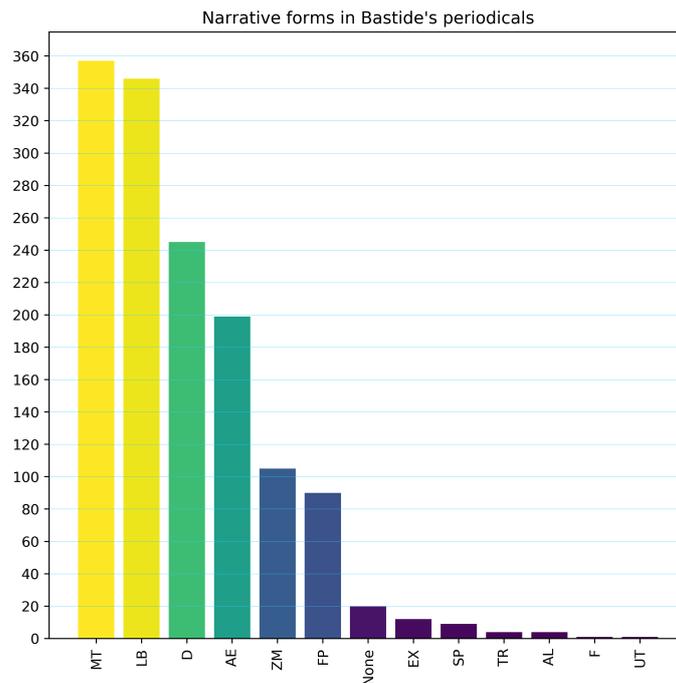


Figure 8: Narrative forms in Bastide's Periodicals.

to the – for the Spectator periodicals – typical multi-layered system of communication (cf. Fischer 2014, 74–83).

Figure 8, which shows a statistical examination of all issues of Bastide's periodicals, further supports the above-mentioned results. It shows that Bastide uses the following narrative forms as predominant communication strategy: metatextual commentaries (MT), letters/letters to the editor (LB), dialogues (D), and general accounts (AE). While the heteroportrait (FD) only stands on fifth position after citation/motto (ZM).

Furthermore, the three bar charts of the topic distribution in Bastide's periodicals (Figure 9, Figure 10, and Figure 11) indicate a wide distribution of topic 9 (virtuous behavior and action) and topic 6 (describing the postulate of enlightened philosophers: '(Self)reflection' (*réflexion*) leading to 'truth' (*vérité*) and knowledge). This focus on virtue and vice is not surprising at all, considering that the Spectator periodicals aim at the moral education of their female and male audience. The readers of the periodicals in general and of Bastide's periodicals in particular are repeatedly exposed to vicious behavior and actions by means of shorter and longer stories in order to guide them to virtuous behavior and actions. A detailed definition or specification of the social norm designated by the term 'virtue', however, is lacking and thus remains undefined; rather, 'being virtuous' is illustrated indirectly through the depiction of its opposite: 'being vicious'. Via the detour of numerous love and relationship stories as well as character portraits which clearly highlight vicious behavior and vicious character traits, the readers are thus led to the desired social norm (cf. Völkl 2022, 291–292).

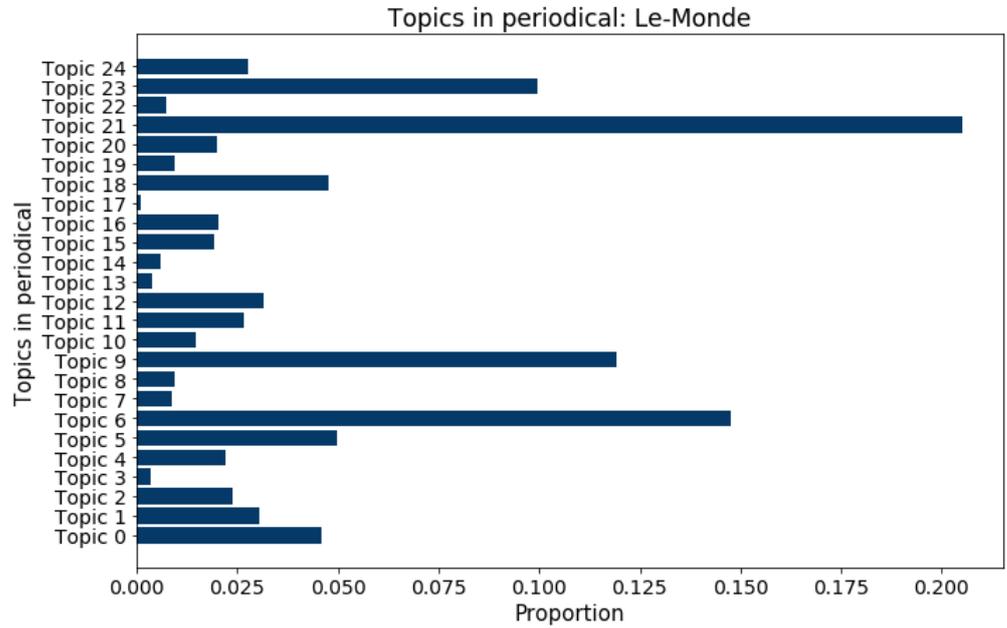


Figure 9: Topics in *Le Monde*.

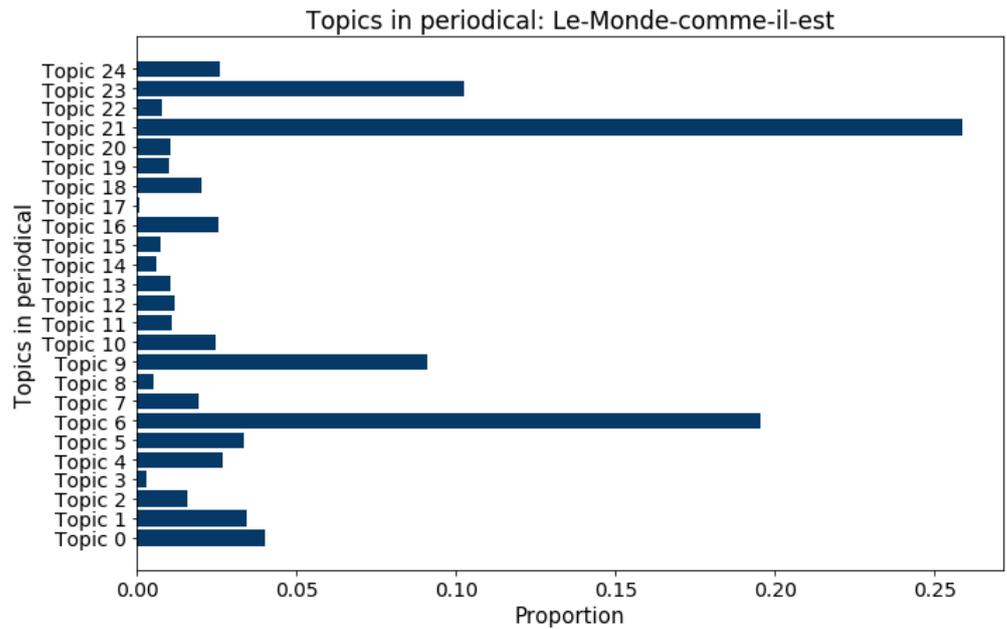


Figure 10: Topics in *Le Monde comme il est*.

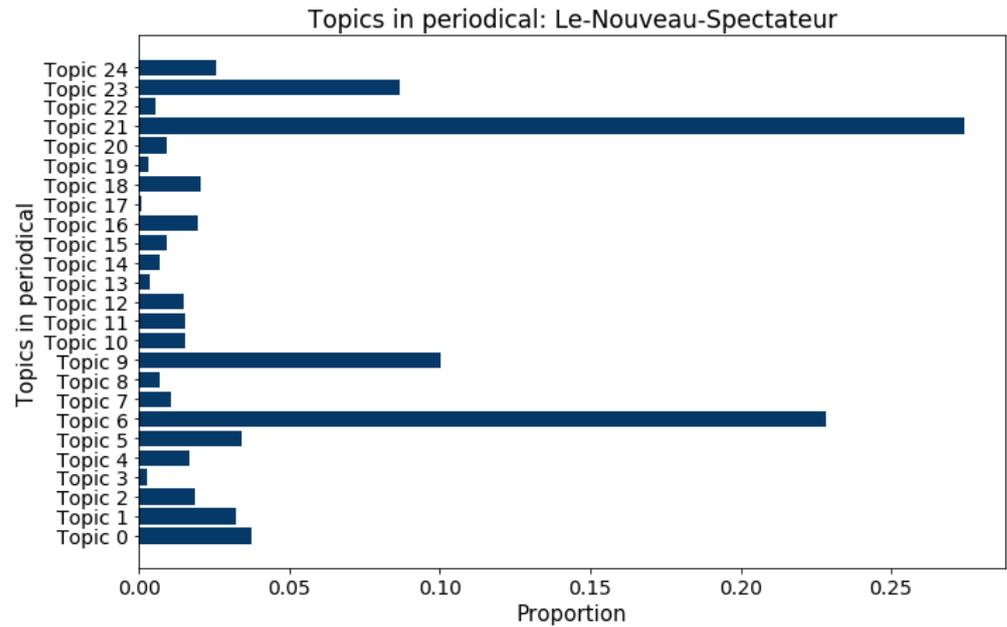


Figure 11: Topics in *Le Nouveau Spectateur*.

6. Topic Modeling in the Spanish-language Spectator Periodicals

Regarding gender-specific topics, the topic modeling results for the Spanish-language periodicals were similar to those of the French-language Spectators. Within the 18 Spanish topics, topics 8, 9, 11 and 17 can be identified as referring to women and men (see Table 4).

Topic 8 is headed by the gender-specific key concept of ‘virtue’ (*virtud*) followed by terms describing elements of a virtuous lifestyle (*vida* – ‘life’, *amor/amar* – ‘(to) love’, *honor/honrar* – ‘(to) honor’), thereby showing considerable similarities to topic 9 of the French-language periodicals.¹⁹ This topic similarity is not surprising at all since the contemporary gender discourse within the French-language periodicals enters the Spanish periodicals – that first appear in Spain from mid-century onward – through numerous translations, imitations, and cultural adaptations. Here, as in France, women and their behavior are held more responsible for the virtuousness of the whole (bourgeois) society than men. More than in other European countries, however, women in Spain are excluded from public life and confined to the private sphere which centers on home, family, and motherhood (cf. Völkl 2022, 229).²⁰ An abundant presence of topic 8 in all Spanish periodicals is thus an expectable development (Figure 12).

Topic 9, with terms such as ‘writing’ (*escribir*), ‘studying’ (*estudiar*), and ‘reading’ (*leer*), alludes to educational activities; the terms ‘science’ (*ciencia*), ‘art’ (*arte*), or ‘history’ (*historia*) of ‘ancient’ (*antiguo*) time to specific study objects. The convergence of these

19. The Spanish topic 8 and the French topic 9 show the following equivalent terms: *virtud* – *vertu*, *vida* – *vie*, *honor/honrar* – *honneur*, *alma* – *ame*, *viciar* – *vice*, *noble* – *noble*, *felicidad* – *bonheur*.

20. Regarding gender discourse in 18th century Spain, see e.g., the monographs and articles by Martín Gaité 1972, Hassauer 1997, Bolufer Peruga 1998, Brink 2008, Capel Martínez 2010, or Gronemann 2013; on the gender discourses in the Spanish novels of the ‘siglo de las luces’, see Hertel-Mesenhöller 2001 or Kilian 2002.

Topic 8	Topic 9	Topic 11	Topic 17
virtud	españa	hijo	mujer
vida	siglo	mujer	moda
amor	lengua	padre	dama
corazon	ciencia	madre	gustar
honor	escribir	criar	hermoso
vivir	nacion	maridar	adornar
placer	mundo	niño	hermosura
alma	estudiar	familia	personar
amar	historia	amor	gracia
mirar	leer	amar	sexo
mundo	libro	edad	figurar
viciar	arte	tratar	señora
honrar	letra	año	bayle
noble	sabio	marido	cortejo
despreciar	idioma	cuidar	mirar
felicidad	españoles	matrimoniar	cabeza
desear	ciencias	señora	naturaleza
efecto	llamar	esposo	rostro
naturaleza	antiguo	hermano	arte
ojo	naciones	cariño	visitar

Table 4: Gender-specific topics in Spanish-language Spectators.

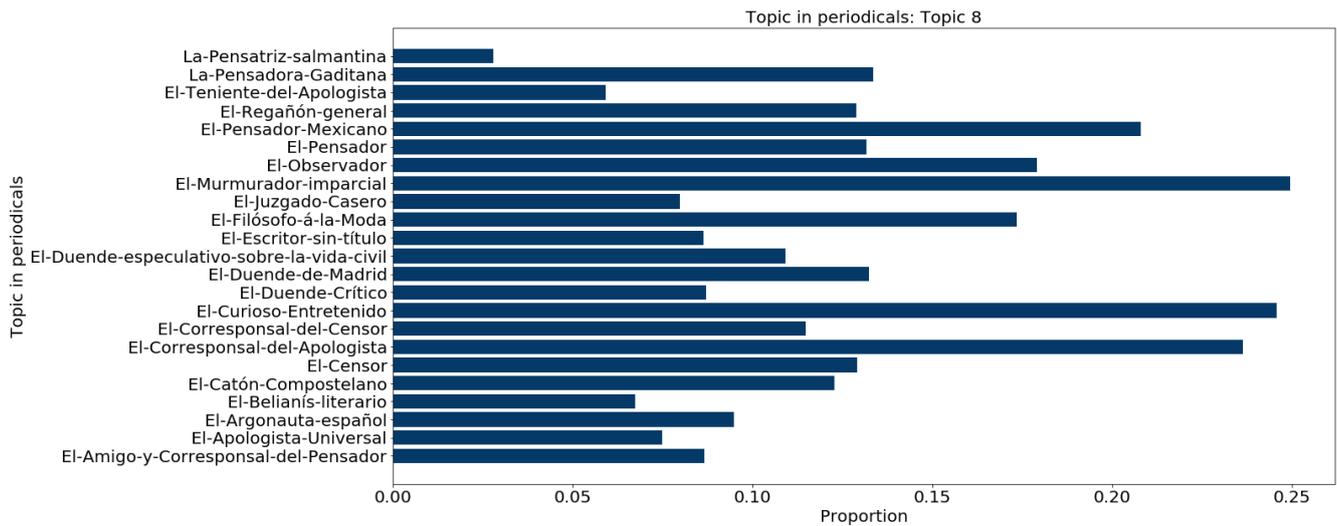


Figure 12: Distribution of topic 8 in Spanish-language periodicals.

terms suggests that this topic describes the education of a bourgeois man, even though no term referring to a male subject (such as *hombre*) – nor to a female subject (such as *mujer*) – can be found. In fact, although the Spanish periodicals grant the female gender a certain capacity for education as well, the terms of topic 9 refer to male formation only. Education for young women is conceived differently to education for young men because (as the French-language periodicals) the Spanish Spectators also propagate a complementary gender model, implying that women and men need to be educated specifically for the correct fulfillment of their gender-specific role in society. The Spanish ideal of the virtuous (bourgeois) woman is also praised in her threefold role as spouse, housewife, and mother, through the fulfillment of which she contributes to the common good of society. This image of woman is conceived in a ‘natural complementarity’ to man, whose ideal image is embodied by the ‘*hombre de bien*’. The latter is characterized by the training of his intellect and subsequently proving useful for his fatherland and the common good. The ‘*hombre de bien*’ of the 18th century is thus not to be confused with the preceding aristocratic ‘*hombre de bien*’ of the 17th century, whose idleness causes his reputation to fall below that of an active citizen – regardless of his status (cf. Heße 2008, 113–130).

Very similar to topic 4 in the French Spectators, the content of **topic 11** supports the construction of the heteronormative society, suggesting the ideal role of ‘woman’ (*mujer*) and man in ‘marriage’ (*maridar, matrimoniar*) where they become ‘mother’ (*madre*) and ‘father’ (*padre*) of ‘children’ (*hijo, niño*). The role of the woman is thus conceived by her ‘husband’s’ (*marido, esposo*) side to whom she is supposed to be a good spouse and housewife. Within the domestic sphere (*familia*), she also receives the role of the ‘caring’ (*cuidar, cariño*) mother who ‘loves’ (*amor, amar*) and ‘raises’ (*criar*) her ‘children’ (*hijo, niño*). Although a rather infrequent topic (Figure 13), it exists throughout all Spanish Spectators.

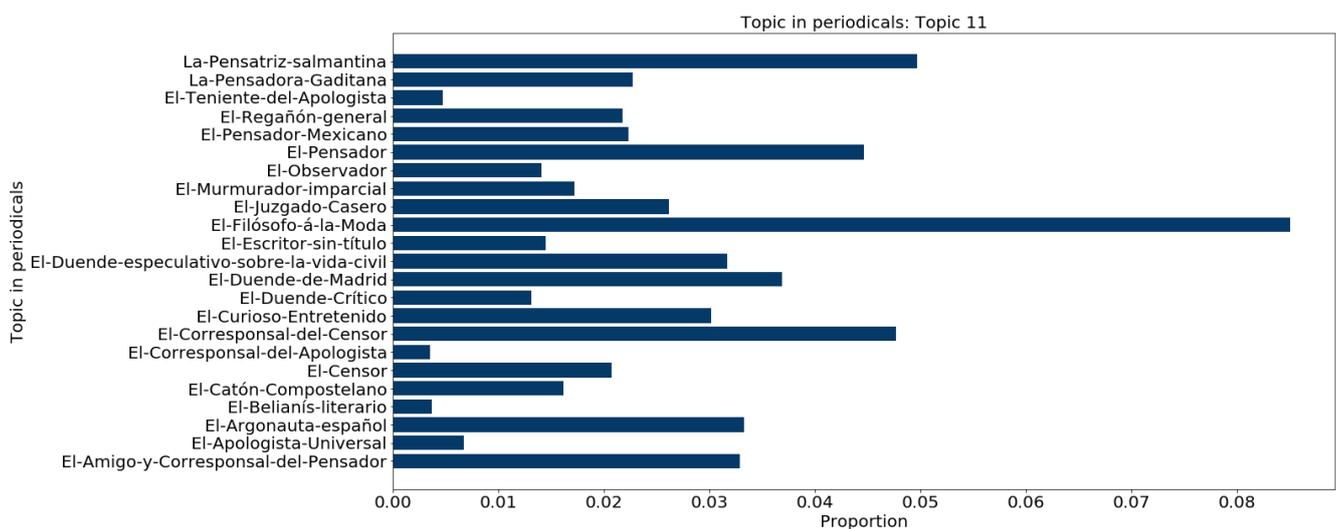


Figure 13: Distribution of topic 11 in Spanish-language periodicals.

Topic 17, represented in Figure 14, points to two discourses associated with the female gender: on the one hand the subject of beauty, on the other hand the then vicious trend of having a relationship with a younger man (*cortejo*). The first eight terms of this topic (*mujer* – ‘woman’, *moda* – ‘fashion’, *dama* – ‘lady’, *gustar* – ‘to please’, *hermoso*

– ‘beautiful’, *adornar* – ‘to adorn’, *hermosura* – ‘beauty’) refer to the semantic field of beauty which pervades the spectatorial gender discourses throughout the century and clearly reflects topic 24 of the French-language periodicals (see Figure 4). In fact, the Spectator periodicals by and large constantly instruct their readers to cultivate external and, increasingly, internal beauty because female beauty is perceived as a pledge for marriage (cf. Schafler 2002, 190) which is seen as the ‘natural’ destiny of the (bourgeois) woman and is thus considered her ultimate goal. At the same time, however, the periodicals warn against falling prey to a cult of beauty that goes hand in hand with the vices of vanity and jealousy. One of these vices, also represented in topic 17, is the gender-stereotypical model of the *cortejo*, i.e., a (younger) man maintaining a very close relationship with a married woman or widow who he ‘visits’ (*visitar*) regularly.²¹ While this (mostly platonic) form of relationship is not a moral offense in aristocratic tradition, it is criticized and stigmatized in the Spectator periodicals.

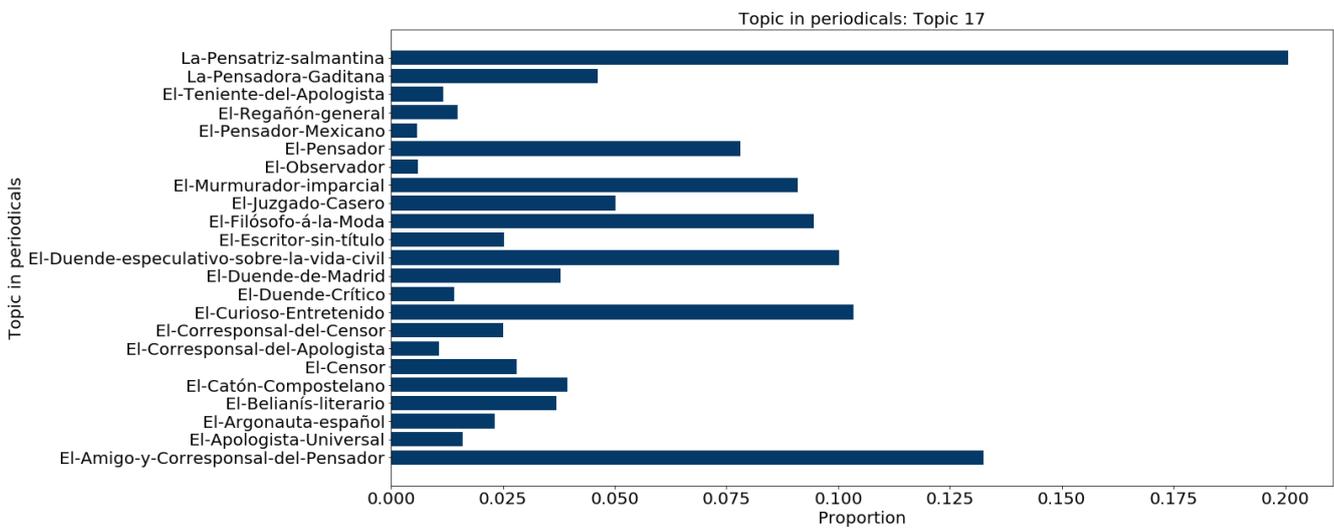


Figure 14: Distribution of topic 17 in Spanish-language periodicals.

Regarding the dissemination of the gender-related topics, the Spanish periodicals pursue a similar strategy to their French-language precursors. Likewise, in the Spanish Spectators virtuous and vicious gender-specific values, norms and practices are mostly conveyed through stories and storytelling. Similar to the French topics 6 and 23, the topic modeling process for the Spanish Spectators revealed topics with a high concentration of narrative vocabulary, such as in topic 2.²² Therein, narrative vocabulary revolves around the semantic fields of movement (e.g., *venir* – ‘to come’, *salir* – ‘to leave’, *llegar* – ‘to arrive’), speech (e.g., *contar* – ‘to narrate’, *entender* – ‘to listen’, *palabra* – ‘word’), and time (e.g., *año* – ‘year’, *hora* – ‘hour’, *noche* – ‘night’), all of which are important components in a story. As can be discerned in Figure 15, topic 2 occurs in all Spanish periodicals.

21. The term ‘cortejo’, which only exists in the masculine form, is not only used to designate the man in this special relationship with a married woman, but also for the woman who allows herself to be courted, and furthermore even to paraphrase the liaison itself (cf. Heße 2008, 135–136).

22. Topic 2: venir salir pasar tomar mano año llegar llamar mil quedar mundo volver entender contar hora amigo ar oír acabar noche palabra

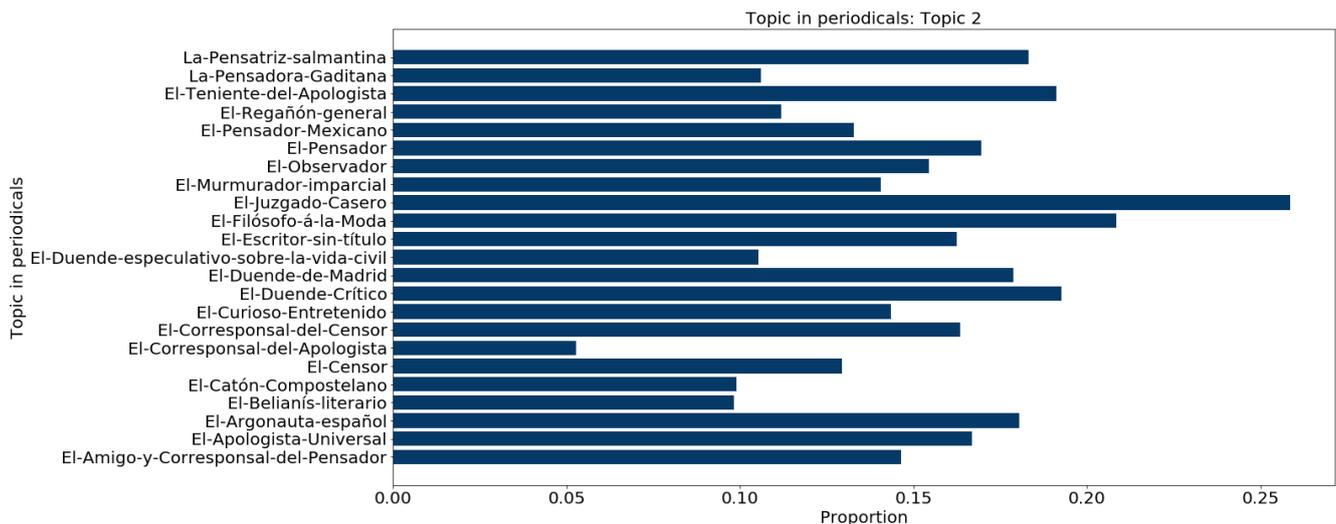


Figure 15: Distribution of topic 2 in Spanish-language periodicals.

7. Conclusion

With their gender-specific discourses, the Spectator press (co-)constructed, preserved, and propagated a bourgeois gender model which is still valid in socio-cultural perception today. This contribution investigates 1,658 French- and 690 Spanish-language issues which were analyzed with topic modeling using LDA. The findings with a focus on gender-specific discourse match and reinforce the results from Völkl's study on narrative and media-specific gender construction within the Spectator periodicals (Völkl 2022).

Using topic modeling, gender-specific topics were identified in the Spectator corpus. Additionally, the application of topic modeling also showed that the Spectator press employed a certain narrative vocabulary (French Spectators: topic 6 and 23; Spanish Spectators: topic 2). Moreover, the comparison between the French- and Spanish-language periodicals rendered similar results: The Spectator corpus of both languages manifested several topics pertaining to a gender-specific discourse. This discourse can be discerned explicitly in topics which exhibit terms referring to female or male stereotypical models, or implicitly in topics which exhibit terms referring to virtuous and vicious gender-specific values, norms, and practices. The study also showed that within the gender-specific topics more terms refer to 'female' than to 'male' (character) traits and behavioral patterns. This result is not surprising neither since the Spectator press supports the creation of a (bourgeois) gender model, based on the complementarity of women and men, and in which the 'female' virtues are assigned a much greater importance for the stability and moral functioning of the society (cf. Pabst 2007; Steinbrügge 1987; Völkl 2022, 120-124). These concordances ascertain that topic modeling as the method used for the present analysis can be successfully employed to question and confirm hypotheses gained through close reading.

In addition to our findings on the gender-specific topics in the Spectators, we described the topic modeling workflow used in DiSpecs in section 4. We aim to make our analysis process transparent for other researchers interested in this method. The research community can also benefit from the primary data available in TEI, and the code which all

are publicly available online (Ertler et al. 2011–2021, Scholger et al. 2019–2021, Scholger et al. 2022).

The DARIAH-DE Notebooks that implement LDA topic modeling proved to be very useful as a basis in our analysis workflow. With some adaptations and additional pre-processing (especially segmentation and lemmatization) and post-processing steps (e.g., results categorization and additional visualizations), we were able to produce comprehensible and insightful results. Our own experience and the comparison with other topic modeling projects allow us to conclude that pre-processing is a crucial part of the analysis since it strongly impacts the quality of the results. The decisions on the respective steps depend on the research material and the specific project goals.

An advantage of topic modeling is the possibility to analyze more content than with close reading, to illustrate the hypothesis on a broader level than through individual examples, and to present the findings using different types of visualizations. Our topic modeling analysis resulted in measurable data of a large text collection's semantic structure which we were able to interpret and comprehensively demonstrate to the Spectators research community. Furthermore, the analysis invoked some new insights into the corpus. Concerning the gender-specific discourse in the Spectators, we saw, e.g., with topic 22 that the French translation of the *Female Spectator* is equipped with a specific semantic vocabulary that can almost exclusively be found in this specific periodical. This result can be attributed to the fact that in this case, we are dealing with a translation and not with a genuine French periodical.

The primary data and the digital scholarly edition also benefit from the topic modeling analysis. With the resulting topics, it is now possible to revise the keywords manually assigned to the individual issues and to further differentiate them. The list of 37 keywords was determined at the beginning of the digital edition project around 2011 and was only minimally expanded in the course of the project. Consequently, the list seems somewhat arbitrary: culture- and language-specific topics – such as 'Apologetic of Spain' which only apply to a few issues – are on the same level as very broad topics such as 'Theatre, Literature, Arts' which combine three areas in one topic. Therefore, the results from topic modeling help to expand and adjust the list of keywords for thematic indexing, thus improving the analysis capabilities within the digital edition as demonstrated in *LdoD Visual* by Portela and Rito Silva 2017. The identified terms in the topics can be incorporated into the TEI metadata header and subsequently used for a more precise and sophisticated search not only at the document level but also on specific text fragments.

Nevertheless, it is necessary to mention certain challenges in using topic modeling. Critics like Dobson argue that the variability of the output depending on the algorithms and set parameters of the method is problematic (cf. Dobson 2021, §20), while Roe, Gladstone, and Morrissey also refer to the probabilistic nature of LDA causing variability in individual runs even with the same parameters (cf. Roe et al. 2016, 4). While we did not compare our LDA results with other algorithms, we agree with Schöch that these variations manifest themselves "in the details of word ranks rather than in the general topics obtained" (Schöch 2017). Parameters have to be tested for individual projects, but once optimized, the method provides relatively stable results.

Furthermore, Murakami et al. as well as Shadrova are skeptical towards methods based on the bag-of-words approach because it ignores the grammatical structures and semantic relations between words (cf. Murakami et al. 2017, 246, Shadrova 2021, 13-14). While we do agree with this statement and believe that every scientific method should be questioned, we also argue that digital methods are not supposed to take on our tasks as humanities experts, but to facilitate research and help us to interpret our data. For these reasons, using a combined approach of topic modeling (and text mining methods in general) and close reading is essential as well as the understanding of the material itself. As Fechner and Weiß point out, it is not the topics that answer research questions themselves, but the researchers through the interpretations of the topics (cf. Fechner and Weiß 2017, 20).

Besides the contribution to the current state of Spectators research and to practical applications of topic modeling, our work also lays the foundations for future work on 18th century literature. The presented results can be compared with similar research on other genres of that time. In addition to the probabilistic topic modeling approach, we intend to integrate transformer-based models to investigate a new corpus of Spanish epistolary novels which are considered to have continued propagating gender-specific values, norms, and practices from the Spectators while also representing an intermediate step towards the 19th century novel.

8. Data Availability

Data can be found here: <https://gams.uni-graz.at/dispecs> and <https://gams.uni-graz.at/spectators>

9. Software Availability

Software can be found here: <https://github.com/distantsspectators>

10. Acknowledgements

We thank the Austrian Academy of Sciences who funded the project DiSpecs, as well as our project team co-members Bernhard Geiger, Christina Glatz, Elisabeth Hobisch, and Philipp Koncar for their cooperation, contribution, and support.

11. Author Contributions

Yvonne Völkl: Results Interpretation, Conceptualization, Writing, Editing

Sanja Sarić: Topic Modeling, Conceptualization, Writing, Editing

Martina Scholger: Project Coordination, Conceptualization, Writing, Editing

References

- Alloghani, Mohamed, Dhiya Al-Jumeily, Jamila Mustafina, Abir Hussain, and Ahmed J. Aljaaf (2020). "Chapter 1: A Systematic Review on Supervised and Unsupervised Machine Learning Algorithms for Data Science". In: *Supervised and Unsupervised Learning for Data Science*. Ed. by Michael W. Berry, Azlinah Mohamed, and Bee Wah Yap. Unsupervised and Semi-Supervised Learning. Springer, 3–22.
- Blei, David M. (2020). "Topic Modeling and Digital Humanities". In: *Journal of Digital Humanities* 1 (2), 8–11. <http://journalofdigitalhumanities.org/2-1/topic-modeling-and-digital-humanities-by-david-m-blei/> (visited on 12/20/2021).
- Blevins, Cameron (2010). "Topic Modeling Martha Ballard's Diary". In: <https://www.cameronblevins.org/posts/topic-modeling-martha-ballards-diary/> (visited on 12/20/2021).
- Bock, Gisela and Margarete Zimmermann, eds. (1997). *Jahrbuch für Frauenforschung. Band 2. Die europäische Querelle des Femmes. Geschlechterdebatten seit dem 15. Jahrhundert*. Metzler.
- Bock, Sina, Keli Du, Michael Huber, Stefan Pernes, and Steffen Pielström (2016). "Der Einsatz quantitativer Textanalyse in den Geisteswissenschaften. Bericht über den Stand der Forschung". In: DARIAH-DE Working Papers (18). Ed. by Mirjam Blümm, Thomas Kollatz, Stefan Schmunk, and Christof Schöch. <http://webdoc.sub.gwdg.de/pub/mon/dariah-de/dwp-2016-18.pdf> (visited on 04/22/2022).
- Bolufer Peruga, Mónica (1998). *Mujeres e Ilustración. La construcción de la feminidad en la Ilustración española*. Inst. Alfons el Magnànim.
- Boyd-Graber, Jordan, Yuening Hu, and David Mimno (2017). "Applications of Topic Models". In: *Foundations and Trends in Information Retrieval*. https://mimno.infosci.cornell.edu/papers/2017_fntir_tm_applications.pdf (visited on 12/20/2021).
- Brink, Margot (2008). "Geschlechterstreit und Dialektik der Aufklärung in Spanien und Frankreich. Die ambivalente Rolle von Vernunft und Natur in Egalitäts- und Komplementaritätstheorien des 18. Jahrhunderts". In: *Heißer Streit und kalte Ordnung. Epochen der 'Querelle des femmes' zwischen Mittelalter und Gegenwart*. Ed. by Friederike Hassauer. Wallstein Verlag, 344–364.
- Capel Martínez, Rosa María (2010). "Prensa y Escritura Femenina en la España Ilustrada". In: *El Argonauta español* (7). <http://journals.openedition.org/argonauta/431> (visited on 12/20/2021).
- Chehal, Dimple, Perul Gupta, and Payal Gulati (2021). "Implementation and comparison of topic modeling techniques based on user reviews in e-commerce recommendations". In: *Journal of Ambient Intelligence and Humanized Computing* (12), 5055–5070. [10.1007/s12652-020-01956-6](https://doi.org/10.1007/s12652-020-01956-6).
- DARIAH-DE (2019). *Notebook Introducing Mallet.ipynb*. <https://github.com/DARIAH-DE/Topics> (visited on 04/01/2019).
- Diaz, Gene (2016). *Stopwords ISO*. <https://github.com/stopwords-iso> (visited on 09/20/2019).
- Dijk, Suzanna van (1988). *Traces de femmes. Présence féminine dans le journalisme français du XVIIIe siècle*. APA, Holland University Press.
- Dobson, James (2021). "Interpretable Outputs: Criteria for Machine Learning in the Humanities". In: *Digital Humanities Quarterly* 2 (15). <http://digitalhumanities.org:8081/dhq/vol/15/2/000555/000555.html> (visited on 12/20/2021).

- Du, Keli (2019). "A Survey On LDA Topic Modeling In Digital Humanities". In: *Book of Abstracts DH2019*. 10.34894/H9UYPI.
- Ertler, Klaus-Dieter (2010). "Die Moralischen Wochenschriften als Vehikel zur diskursiven Ausdifferenzierung der Nation in Spanien". In: *Beiträge zur Nationalisierung der Kultur im Spanien des aufgeklärten Absolutismus*. Ed. by Jan-Henrik Witthaus. Peter Lang, 93–107.
- Ertler, Klaus-Dieter, Alexandra Fuchs, Michaela Fischer-Pernkopf, Elisabeth Hobisch, Martina Scholger, and Yvonne Völkl (2011–2021). *The Spectators in the International Context*. <https://gams.uni-graz.at/spectators> (visited on 12/20/2021).
- Fechner, Martin and Andreas Weiß (2017). "Einsatz von Topic Modeling in den Geschichtswissenschaften: Wissensbestände des 19. Jahrhunderts". In: *Zeitschrift für digitale Geisteswissenschaften*. 10.17175/2017_005.
- Fischer, Michaela (2014). *Die Figur des Lesers im Kommunikationssystem der Spectateurs*. Peter Lang.
- Fischer-Pernkopf, Michaela, Veronika Mussner, and Klaus-Dieter Ertler (2018). *Die «Spectators» in Frankreich. «Le Nouveau Spectateur» und «Le Monde comme il est» von Jean-François de Bastide*. Peter Lang.
- Gephi.org (2008–2021). *Gephi V.0.9.2*. <https://gephi.org/> (visited on 12/20/2021).
- Gronemann, Claudia (2013). *Polyphone Aufklärung. Zur Textualität und Performativität der spanischen Geschlechterdebatten im 18. Jahrhundert*. Vervuert.
- Hassauer, Friederike (1997). "Die Seele ist nicht Mann, nicht Weib. Stationen der Querelles des Femmes in Spanien und Lateinamerika vom 16. zum 18. Jahrhundert". In: *Die europäische Querelle des Femmes. Geschlechterdebatten seit dem 15. Jahrhundert*. Ed. by Gisela Bock and Margarete Zimmermann. Jahrbuch für Frauenforschung. Metzler, 203–238.
- Hertel-Mesenhöller, Heike (2001). *Das Bild der Frau im spanischen Roman des 18. Jahrhunderts. Im Spannungsfeld von Lebenswirklichkeit und Fiktion*. Vervuert.
- Heße, Kristina (2008). *Männlichkeiten im Spanien der Aufklärung. Der Diskurs der moralischen Wochenschriften 'El Pensador', 'La Pensadora gaditana' und 'El Censor'*. Logos.
- Hobisch, Elisabeth (2017). *La forma epistolar en los espectadores españoles: Características y tipología de las cartas*. Peter Lang.
- (2018). "Les stratégies publicitaires dans les lettres des 'spectateurs' espagnols". In: *Discourses on Economy in the Spectators / Discours sur l'économie dans les spectateurs*. Ed. by Klaus-Dieter Ertler, Samuel Baudry, and Yvonne Völkl. Dr. Kovač, 199–214.
- Honegger, Claudia (2011). "Die kognitiven Prinzipien der neuen Wissenschaften vom Menschen und die Genese einer weiblichen Sonderanthropologie in Frankreich". In: *Die gesellschaftliche Verortung des Geschlechts. Diskurse der Differenz in der deutschen und französischen Soziologie um 1900*. Ed. by Theresa Wobbe. Campus, 93–113.
- Horstmann, Jan and Rabea Kleymann (2019). "Alte Fragen, neue Methoden – Philologische und digitale Verfahren im Dialog. Ein Beitrag zum Forschungsdiskurs um Entsagung und Ironie bei Goethe". In: *Zeitschrift für digitale Geisteswissenschaften*. 10.17175/2019_007.
- Hu, Yuening, Jordan Boyd-Graber, Brianna Satinoff, and Alison Smith (2014). "Interactive topic modeling". In: *Mach Learn* (95), 423–469. 10.1007/s10994-013-5413-0.
- Jacobs, Eva, W. H. Barber, Jean H. Bloch, F. W. Leakey, and Eileen Le Breton (1979). *Woman and society in eighteenth-century France : essays in honour of John Stephenson Spink*. Athlone Press.

- Jockers, Matthew Lee (2014). *Text analysis with R for students of literature*. Quantitative Methods in the Humanities and Social Sciences. Springer, Cham.
- Kilian, Elena (2002). *Bildung, Tugend, Nützlichkeit. Geschlechterentwürfe im spanischen Aufklärungsroman des späten 18. Jahrhunderts*. Königshausen & Neumann.
- Krautter, Benjamin, Janis Pagel, Nils Reiter, and Marcus Willand (2020). "»[E]in Vater, dächte ich, ist doch immer ein Vater«. Figurentypen im Drama und ihre Operationalisierung." In: *Zeitschrift für digitale Geisteswissenschaften*. 10.17175/2020_007.
- Laqueur, Thomas (2003[1990]). *Making Sex. Body and Gender from the Greeks to Freud*. Harvard University Press.
- Liu, Alan, Scott Kleinman, Jeremy Douglass, Thomas Lindsay, Ashley Champagne, and Jamal Russell (2017). "Open, Shareable, Reproducible Workflows for the Digital Humanities: The Case of the 4Humanities.org 'WhatEvery1Says' Project". In: *Digital Humanities 2017. Conference Abstracts*. McGill University and Université de Montréal, 95–98. <https://dh2017.adho.org/abstracts/DH2017-abstracts.pdf> (visited on 11/20/2021).
- Marjanen, Jani, Elaine Zosa, Simon Hengchen, Lidia Pivovarova, and Mikko Tolonen (2020). "Topic modelling discourse dynamics in historical newspapers". In: *arXiv preprint*. 10.48550/arXiv.2011.10428.
- Martín Gaité, Carmen (1972). *Usos amorosos del dieciocho en España*. Siglo XXI de España Editores.
- Mauzi, Robert (1969). *L'idée du bonheur dans la littérature et la pensée françaises au XVIIIe siècle*. Colin.
- McCallum, Andrew Kachites (2002–2018). *MALLET. A Machine Learning for Language Toolkit*. V. 2.0.8. <https://mimno.github.io/Mallet/index> (visited on 11/28/2021).
- Murakami, Akira, Paul Thompson, Susan Hunston, and Dominik Vajn (2017). "What is this corpus about?: using topic modelling to explore a specialised corpus". In: *Corpora* 2 (12), 243–277. 10.3366/cor.2017.0118.
- Mussner, Veronika (2016). "Die Moralischen Wochenschriften in Frankreich. 'Le Monde comme il est' von Jean-François de Bastide als Spiegel seiner Zeit". MA thesis. Universität Graz.
- Nelson, Robert K. (2020). *Mining the Dispatch*. <https://dsl.richmond.edu/dispatch/> (visited on 12/20/2021).
- Norton, Brian Michael (2015). "The Spectator, Aesthetic Experience and the Modern Idea of Happiness". In: *English Literature* 2 (1), 87–104.
- Pabst, Esther S. (2007). *Die Erfindung der weiblichen Tugend. Kulturelle Sinngebung und Selbstreflexion im französischen Briefroman des 18. Jahrhunderts*. Wallstein.
- Portela, Manuel and António Rito Silva (2017). *Arquivo LdoD: Arquivo Digital Colaborativo do Livro do Desassossego*. <https://ldod.uc.pt/>.
- Roe, Glenn, Clovis Gladstone, and Robert Morrissey (2016). "Discourses and Disciplines in the Enlightenment: Topic Modeling the French Encyclopédie". In: *Frontiers in Digital Humanities* (2). 10.3389/fdigh.2015.00008.
- Schaufler, Birgit (2002). *'Schöne Frauen – starke Männer'. Zur Konstruktion von Leib, Körper und Geschlecht*. Leske + Budrich.
- Schöch, Christof (2016). "Topic Modeling with MALLET. Hyperparameter Optimization". In: <https://dragonfly.hypotheses.org/1051> (visited on 12/20/2021).

- 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 12/20/2021).
- Scholger, Martina, Bernhard Geiger, Elisabeth Hobisch, Philipp Koncar, Sanja Sarić, Yvonne Völkl, and Christina Glatz (2022). *Distant Spectators. Distant Reading for Periodicals of the Enlightenment (DiSpecs)*. GitHub repository. <https://github.com/distant-spectators/DiSpecs>.
- (2019–2021). *Distant Spectators. Distant Reading for Periodicals of the Enlightenment (DiSpecs)*. <https://gams.uni-graz.at/dispecs> (visited on 12/20/2021).
- 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).
- Sieuzac, Laurence (2009). "Éducation et vocation de la femme au Siècle des lumières". In: *Genre & Éducation. Former, se former, être formée au féminin*. Ed. by Paul Pasteur et al. Presses universitaires de Rouen et du Havre, 271–287.
- Steinbrügge, Lieselotte (1987). *Das moralische Geschlecht. Theorien und literarische Entwürfe über die Natur der Frau in der französischen Aufklärung*. Beltz.
- Sullerot, Evelyne (1966). *Histoire de la Presse féminine en France des origines à 1848*. Armand Colin.
- TEI Consortium (2021). *Guidelines for Electronic Text Encoding and Interchange*. <http://www.tei-c.org/P5> (visited on 12/20/2021).
- Völkl, Yvonne (2022). *Spectatoriale Geschlechterkonstruktionen: Geschlechtsspezifische Wissens- und Welterzeugung in den französisch- und spanischsprachigen Moralischen Wochen-schriften des 18. Jahrhunderts*. transcript.