

Applying Topic Modeling to Art History Articles: An Analysis of the Journal of the Brazilian Service/Directorate of Historic and Artistic Heritage (1937-1961)

Abstract

This paper addresses a critical gap in digital art history by employing Natural Language Processing to analyze the Journal of the Brazilian Service/Directorate of Historic and Artistic Heritage from 1937 to 1961. Using a topic modeling method, namely *Structural Topic Modeling*, we reveal representative themes, authors, and articles in this influential journal shaping Brazilian art history. Our analysis not only confirms the emphasis on material heritage with European connections but also highlights topics that have not received due attention to date, including contributions from other civilizational matrices like the African and Indigenous ones. This study showcases the potential of computational textual methods, particularly topic modeling, for providing valuable insights in art historical research and emphasizes the importance of computational approaches in investigating cultural heritage and scholarly discourse.

Keywords

Art History. Computational Methods of Analysis. Natural Language Processing. Topic Modeling.

1. Introduction

The past decade has seen tremendous growth in the use of digital resources, methods, and tools in the art history, leading to the constitution of an authentic sub-field of the discipline that is usually referred to as “digital art history” [1, 2]. The most common methods of analysis in digital art history are shared with other branches of the so-called Digital Humanities and include: spatial analysis; network analysis; image analysis; and textual analysis [3]. Of all these methods, however, the ones linked to textual analysis remain persistently underrepresented in conferences, journals, and edited volumes dedicated to digital art history [1, p. 7].

This paper seeks to contribute to filling this gap using methods related to Natural Language Processing (NLP). More specifically, we apply a topic modeling method in order to reveal the most representative themes, authors and articles, as well as temporal thematic trends, within a seminal journal for the constitution of Brazilian art history: the *Revista do Serviço do Patrimônio Histórico e Artístico Nacional* (Journal of the National Service of Historic and Artistic Heritage), created in 1937 and renamed simply *Revista do Patrimônio Histórico e Artístico Nacional* (Journal of the National Historic and Artistic Heritage) in 1946. We consider here the full textual content of the first fifteen issues of the journal, published between 1937 and 1961 (see Figure 1). In addition, we verify the relevance of the results obtained by comparing them with a significant number of studies dealing with the editorial line of the journal [4, 5, 6, 7, 8].

CHR 2023: Computational Humanities Research Conference, December 6 – 8, 2023, Paris, France



© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

In the critical fortune of the journal, there is a widespread consensus that sees it as a vehicle that affirmed a relatively narrow conception of Brazilian heritage. In a nutshell, this conception emphasized the material nature of monuments and art works deemed valuable, its connection to European (more specifically Portuguese) cultures, and its location in important – but few – administrative centers of colonial Brazil. The critical consensus regarding the journal’s editorial emphasis has been articulated by Brazilian scholars since at least the 1990s. For example, in her seminal work on the genesis of heritage preservation practices in Brazil, Marcia Chuva ([9], p. 230) characterized the journal of the Service of Historic and Artistic Heritage as “a periodical specialized in the ‘history of material civilization in Brazil’, with a temporal concentration on the Portuguese colonial period”. Also according to Chuva, the concept of “material civilization in Brazil” was formulated by the jurist, politician, professor and writer Afonso Arinos de Melo Franco, who had a close relationship with service/directorate. In special, Arinos taught a course for the service’s employees entitled “*Desenvolvimento da Civilização Material no Brasil*” (Development of Material Civilization in Brazil), which resulted in a volume published by the institution in 1944 [10]. In essence, the analysis reported in this paper reiterates this consensus.

Usually, in art history, if computational methods merely confirm the understanding of certain phenomena already sketched out by previous research, they are considered to be of little use and are little appreciated. However, it is worth remembering that this demand for “new” results is much less pressing in other disciplines. For example, “Pierre Bourdieu reminded us that to confirm what was intuitively formulated is indeed to advance research, both through the development of a more objective method, and through the greater finesse of the results acquired” [11, p. 8]. Moreover, we would like to stress that the results of our analysis do not merely confirm the consensus developed earlier, but also do indicate some topics which, if not completely ignored in the critical reception of the journal, have not received due attention to date. Indeed, our analysis reveals that articles published in the journal addressed themes that may be unexpected to several art historians, notably other civilizational matrices that contributed to the formation of Brazil, such as the African and specially the Indigenous ones.

Beyond their originality regarding the discipline of art history, we believe that the NLP methods we employ potentially transcend our specific object of study. Indeed, topic modeling methods are already used in the analysis of other textual corpora, concerning, for example, electoral campaigns, political speeches, or the influence of actors in social networks [12, 13, 14]. Therefore, if these methods are effective in achieving the proposed objective, they could be applied to other journals in the humanities that do not have an established critical fortune, and that typically employ a variety of formatting styles and do not adhere to standard formats such as those found in natural and social sciences journals [15].

The remainder of the paper is organized as follows. Section 2 provides a brief overview of the inception and trajectory of the journal between 1937 and 1961, discusses its main physical characteristics and points to the irregularity in the publication of its issues. In Section 3, we introduce the methodological framework employed in the textual analysis, elucidating the principles and steps involved in topic modeling method employed, namely Structural Topic Modeling. We also discuss the rationale behind selecting its parameters aimed at achieving optimal accuracy. Next, Section 4 presents the results of our automated topic analysis of the journal’s articles. We outline the prevalent and less prevalent topics that emerged and highlight specific themes, which provide valuable insights into the journal’s content. Lastly, in Section



Figure 1: Facsimiles of the covers of the issues of the journal of the Service/Directorate of Historic and Artistic Heritage published between 1937 and 1961.

5, we summarize our main findings, address the limitations of the automated approach, and emphasize the potential for future research.

2. The Journal of the SPHAN/DPHAN, 1937-1961: A Survey

SPHAN – Serviço do Patrimônio Histórico e Artístico Nacional (National Service of Historic and Artistic Heritage) was the first designation of the Brazilian federal agency for the protection of cultural heritage, which dates back to 1936. During its almost nine decades of operation, the institution has been renamed several times: between 1946 and 1970, for example, it was called *DPHAN – Diretoria do Patrimônio Histórico e Artístico Nacional* (National Directorate of Historic and Artistic Heritage). Nowadays, it still exists, under the name *IPHAN – Instituto do Patrimônio Histórico e Artístico Nacional* (National Institute of Historic and Artistic Heritage).

The SPHAN began to function in 1936, linked to the Brazilian Ministry of Education and Public Health, directed by Gustavo Capanema; however, only in the following year it was officially created, with the enactment of the Law nº 378 on January 13, 1937 [16]. This law defined SPHAN’s objectives as follows: “the National Institute of Historic and Artistic Heritage is hereby created, with the purpose of promoting, throughout the country and on a permanent basis, the *tombamento* (registration in a *Livro do Tombo*¹), conservation, improvement and knowledge of the national historic and artistic heritage” [18].

One of the first actions aimed at achieving these objectives was the creation of SPHAN’s publications sector in 1937 [4, 7]. In its early years, this sector was composed of two main elements: (a) a journal which, in each issue, brought together works by diverse authors (articles, essays, technical studies, reproductions or transcriptions of historical sources, etc); and (b) a

¹The term *tombamento* (verb: *tomb*) is deeply rooted in Luso-Brazilian history. Originating in the Latin word for archive or repository *tumulum*, the term *tombamento* was historically associated with official registries of property and wealth. [However] When the term *tombamento* was integrated into Brazilian preservationist law in the mid-1930s, its meaning was fixed in an administrative process of formally inscribing important historical sites and works of art in official registries, known as *Livros do Tombo* [17].

series of larger monographs entitled *Publicações* (literally “Publications”), each authored by a single author, which discussed specific monuments and artists, but also included historical studies, guides, catalogues of collections, etc. [7]. During the period considered here, both the journal and the *Publicações* series had a single director: the lawyer, journalist and writer Rodrigo Melo Franco de Andrade, who also served as director of SPHAN/DPHAN between 1937 and 1967. Andrade played a central role in the configuration of the journal’s editorial policy, by directly choosing the authors who published in it, as well as in the writing of several articles, delimitating their themes, proposing adjustments and corrections, and working as a translator.

Arguably, the journal of the SPHAN/DPHAN was the first Brazilian publication to deal exclusively with historic/artistic monuments and related themes. In the “Program” that opened its first issue, Andrade defined the *raison d’être* of the journal in terms in tune with those of the Law n° 378: “The aim here is above all to disseminate knowledge of the values of art and history that Brazil possesses and to make a committed contribution to their study” [19, not paged]. There were previous efforts in this direction in Brazil, but, as Andrade pondered, these were “scattered in pamphlets, newspapers and magazines, whose search requires effort and patience” [19, not paged]. In this sense, the journal sought to assert itself as a locus that would centralize discussions on Brazilian historic and artistic heritage. The works published there routinely made use of what Luciano dos Santos Teixeira dubbed “protocols of truth” [8, not paged] associated with the idea of scientific objectivity, such as: the use of historical sources and their transcriptions, understanding them as arguments of authority; the narrative and descriptive character of the texts; the use of instruments of erudition and academic seriousness, such as bibliographies, notes, indexes, etc.

The journal’s austere graphic presentation seemed to reiterate its pretensions of scientific objectivity. Between 1937 and 1961, it maintained constant physical characteristics: “book-like binding, with dimensions of 17.5 x 23.5 cm; an average of approximately 322 pages per issue and many illustrations. The text pages were printed on offset paper and the illustrations were printed on high quality *couché* paper [...] there were no advertisements or announcements of any kind in the pages of the journal” [5, p. 73]. During its publication from 1937 to 1961, the journal featured articles signed by diverse authors and eight anonymous articles².

The journal’s publication schedule exhibited irregularity. Based on the dates printed on its covers, the eleven initial issues were released annually between 1937 and 1947. Subsequently, a hiatus occurred, and the last issues that we consider were dated 1955, 1956, 1959, and 1961. The analysis of the actual printing dates reveals, however, that in general there was “a gap between the year attributed to an issue, and the year of its actual circulation among readers” [7, p. 82]. At first, this lag averaged one year, but increased in the mid-1940s. For example, issues dated on the cover 1944 and 1947 did not circulate until 1947 and 1954, respectively. The irregularity in the effective circulation of the journal points to its close link with the cultural policy promoted by Brazilian President Getúlio Vargas, who ruled for the first time between 1930 and 1945, and again between 1951 and 1954. With Vargas’ departure from presidency in 1945, “it became more difficult to obtain funds for the printing of the journal – which explains why only one volume was published until 1950, the time interval between one Vargas government and another” [7, p. 86]. We believe it is important to keep these irregularities in mind when analyzing, for

²All these eight anonymous articles appeared in the first issue of the journal.

example, how the different topics appear throughout the time period considered in the paper.

3. Overview of the Topic Modeling Method

The computational topic modeling method employed to analyze the selected articles is called *Structural Topic Modeling (STM)*, which is widely used in the field of textual analysis in connection with social sciences as an aid in linguistic, political and psychological analysis [13, 14, 20]. This method aims to identify the main themes present in a corpus, as well as the patterns and trends related to the broader scope of the issues addressed. It assumes that the corpus can be seen as a sample of a statistical generative model in which each text is generated by successively sampling words taking into account their estimated links with certain themes, without direct semantic or syntactic considerations [12]. The underlying assumption of this hypothesis is twofold by considering that each text is composed of a weighted mixture of topics, and that each word composing it is associated with one of these topics.

In general terms, a topic is defined as a set of semantically related words within a specific theme, which recur in one or more texts. The relative importance of each vocabulary word in the set of texts with a topic expresses the degree of the word's semantic link to the topic in question. Therefore, the words with the highest relative importance in a topic determine the specific theme contemplated. As a result of the analysis, the relative importance of each topic in each text is obtained, as well as the relative importance of each word in each topic. Thus, the representation of each text in the form of a mixture of topics reveals associations between the information contained in it that do not appear immediately [20, 21].

A crucial parameter of the method is the number of topics, which must be specified beforehand. The selection of an appropriate number of topics plays a central role in the analysis, as it affects the granularity and interpretability of the results. Too few topics may lead to oversimplification, while an excessive number of topics can result in ambiguity and hinder the identification of meaningful patterns. Therefore, empirical preliminary experiments are usually conducted on a subset of the corpus to determine the number of topics that best capture the underlying themes.

One of the distinguishing features of the STM method employed in this study is its ability to explore groupings of texts based on shared metadata, allowing for a more specific analysis of each group [14]. By assuming that texts within a single group exhibit certain similarities, the STM method incorporates these groupings as additional information within the model. This is achieved by considering text-level covariates that can influence both the prevalence and content of the topics. To initiate the analysis with STM, one must first identify the metadata that will be used to define the groupings. In our case, we utilized specific characteristics of the articles, such as the author's name and the publication year, as the metadata for defining article groupings based on shared values. The underlying assumption is that texts authored by the same author may share a similar writing style and line of thinking. Similarly, articles appearing in the same journal issue may exhibit thematic and vocabulary similarities.

4. Computational Experiments and Analysis

4.1. Collecting the Articles and Preparing the Corpus

The issues of the journal of the SPHAN/DPHAN are accessible for individual download in pdf format from the official IPHAN's website³. For our analysis, we specifically collected the pdf files corresponding to the fifteen journal issues published between 1937 and 1961 from the website. These issues, comprising over 4700 pages in total, were not originally created in a digital format; instead, they exist as collections of images of the original printings that were digitized to create the pdf files. To extract the contents from these images, we utilized Optical Character Recognition (OCR) processing. The OCR program converted the scanned images into machine-readable texts encoded using the UTF8 encoding. Subsequently, the resulting converted files were divided, ensuring that each of the 151 articles was accommodated in an individual binary file, each encoded in UTF8. It is important to acknowledge that, during the OCR process, a few errors may have arisen due to word splitting or non-standard page formats. However if we consider the total length of the corpus and the particular behavior of the method employed (see Section 3), the impact of these errors is negligible.

Table 1 is a partial view of the corpus constructed as described above, where a selection of authors (out of a total of 73 authors) is presented along with the respective journal issues in which their articles have been published. The criteria for inclusion in the table are based on their contributions to articles that meet a specific threshold of relevance within the topic analysis conducted in our experiments. Specifically, authors are included in the table if they have authored articles that are deemed relevant to a particular topic according to the outcomes of the topic analysis, thus showcasing a meaningful association with the identified topic.

Prior to subjecting the articles to topic analysis, a preprocessing phase was conducted. Standard techniques were employed to eliminate accents, punctuation, and stopwords. Additionally, singularization and 2-gram detection methods were also applied to further refine the text data. The resulting dataset consisted of 4526 tokens.

4.2. Summary of Results

Our analysis of the set of articles from the journal of the SPHAN/DPHAN employed the STM method, implemented in the programming language R through the *quanteda* and *stm* packages [22]. We selected the number of 15 topics based on empirical experiments, which demonstrated its ability to produce the most accurate results. Figure 2 presents a summary of the findings in which each topic is identified by a list of the most relevant words characterizing it. The topics are arranged in descending order of average relevance in the set of articles, providing a quick visualization of the main themes addressed in the journal and their relative importance. As an overall analysis, it is worth noting that the average relevance of topics ranges from 0.0194 to 0.1176, with an average of 0.0667 and a standard deviation of 0.0262. Few topics have average relevance that falls beyond one standard deviation from the mean. Topics 6 and 9 are the most relevant, indicating that a significant number of articles are influenced by them. On the other

³URL: <http://portal.iphan.gov.br/publicacoes/lista?categoria=23&busca>.

Table 1
Selected authors and corresponding journal issues.

Author	Issues	Author	Issues	Author	Issues
Alberto Lamego (AlbL)	1938, '40	Francisco Marques dos Santos (FMdS)	1937, '38, '39, '41	Lucio Costa (LcCs)	1937, '39, '41
Alfredo Galvao (AlfG)	1956, '59, '61	Gastao Cruls (GstC)	1941, '42	Luiz Jardim (LzJr)	1939, '40
Aluizio de Almeida (AlaA)	1945	Gilberto Ferrez (GlbF)	1946	Luiz Saia (LuzS)	1939, '44
Artur Cezar Ferreira Reis (ACFR)	1941, '42, '43, '44, '46, '47	Gilberto Freyre (GlbF)	1937, '43	Manuel Bandeira (MnlB)	1938, '42
Augusto de Lima Junior (AdLJ)	1938, '45	Hannah Levy (HnnL)	1940, '41, '42, '44, '45	Maria de Lourdes Pontual (MdLP)	1940
Ayrton Carvalho (AyrC)	1942	Helcia Dias (HlCd)	1939	Nair Batista (NrBt)	1939, '40, '41
Bonifacio Jansen (BnfJ)	1955	Heloisa Alberto Torres (HIAT)	1937	Noronha Santos (NrnS)	1937, '40, '42, '44, '46, '47
Carlos Estevaso (CrIE)	1938, '39	J. Moritz Rugendas (J.MR)	1956	Paulo Thedim Barreto (PITB)	1937, '38, '47
Carlos Ott (CrIO)	1943, '47, '56, '59, '61	Joao Miguel dos Santos Simoes (JMdSS)	1959	Raimundo Lopes (RmnL)	1937, '38
Carlos Tasso de Saxe Coburgo e Braganca (CTdSCeB)	1961	Joaquim de Souza Leao (JdSL)	1945, '46, '56	Raimundo Trindade (RmnT)	1943, '44, '45, '55, '56, '59
Clemente Maria da Silva Nigra (CMdSN)	1941, '42, '43, '44, '45	Jose Antonio Gonsalves de Mello (JAGdM)	1961	Rodrigo M. F. de Andrade (RMFdA)	1937, '38
Curt Nimuendaju (CrtN)	1944	Jose de Almeida Santos (JdAS)	1942	Salomao de Vasconcelos (SldV)	1938, '39, '40, '41, '45, '55
David James (DvdJ)	1955, '56	Jose de Araujo Wanderley Pinho (JdAWP)	1940	Serafim Leite (SrfL)	1942, '44
E Orosco (EOrs)	1941	Jose de Souza Reis (JdSR)	1939, '55	Sergio Buarque de Holanda (SBdH)	1941
Epaminondas de Macedo (EpdM)	1937	Judite Martins (JdtM)	1939, '40, '61	Sylvio de Vasconcellos (SydV)	1959
Estevaso Pinto (EstP)	1938, '43	Lourenco Luis Lacombe (LrLL)	1938, '44	Venancio Willeke (VnnW)	1956

hand, topics 5 and 15 are the least relevant, showing their relatively small appearance in the articles. The remaining topics fall between 0.0445 and 0.0879.

4.3. Discussion of the Results

A comprehensive analysis of the detected topics reveals their alignment with the consensus mentioned in the Introduction. More specifically, the topic analysis largely reiterates the journal's characterization advanced by Chuva and other scholars [9]. Even if the temporal focus is vaguely mentioned (e.g., the term *antigo* – ancient – in topics 3 and 15), topic 6 clearly establishes the connection of the journal's profile with Portugal as a pivotal characteristic. On the other hand, the essentially “material” nature of the heritage discussed in the journal is evident through various terms found in many topics, such as *agua*, *ceramica*, *madeira*, and *pedra* (water, ceramics, wood, and stone, respectively). As shown in Figure 3, this is further supported by the main artistic production modalities discussed in the journal, such as *azulejo*, *desenho*, *fotografia*, *gravura*, and *pintura* (tiles, drawing, photograph, engraving, and painting, respectively) detected by topics 6 and 9.

Figure 3a illustrates the behavior of these two topics based on the expected proportion of each topic in selected articles where they are most relevant. Only one article is selected per year. The y-axis represents the publication years of the journal issues, while the x-axis denotes the expected proportion of the respective topic in the selected articles. Accordingly, topics 6 and 9 are most relevant to articles authored by Augusto de Lima Junior (AdLJ) and David James (DvdJ), respectively, with the highest estimated relevance of the topics observed in the 1938 and

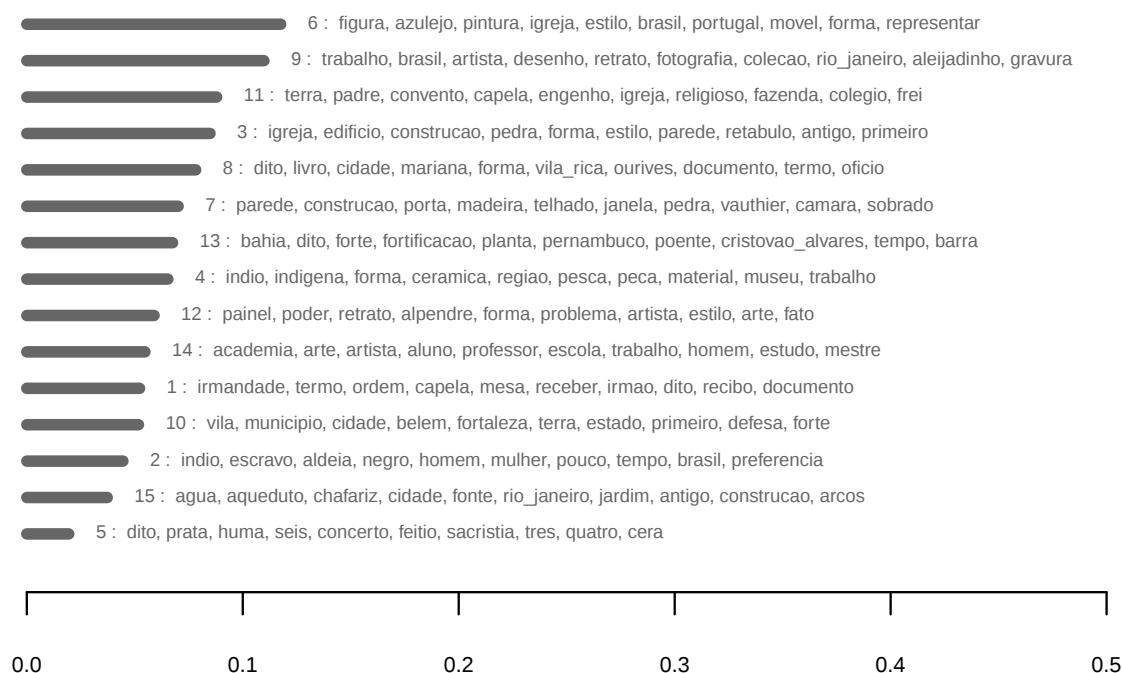


Figure 2: Synthesis of the topics found in the articles of the journal of the SPHAN/DPHAN published between 1937 and 1961.

1955 issues. There are no articles in the issues that do not appear in the plot where the selected topics are most relevant.

Figure 3b provides a graphical representation of the estimated proportion of the same topics across the journal issues. It enables a visual analysis of the trends in topic prevalence within all articles (not only those shown in Figure 3a). The x-axis of the plot represents the publication years of the journal issues, while the y-axis represents the estimated proportion of the topic. The curve shown in the plot represents the trend of the topic proportion along the years, obtained with a regression method on the estimated proportions. For example, if the plot shows an upward-sloping line, it indicates that as the metadata variable increases, the estimated proportion of the topic also tends to increase. Conversely, a downward-sloping line suggests that as the metadata variable increases, the topic proportion decreases.

Still linked to the idea of a Brazilian “material civilization”, a number of terms directly associated with the idea of *construcao* – i.e, construction in an architectural sense – stand out. These terms include architectural typologies, such as *aqueduto*, *capela*, *chafariz*, *convento*, and *igreja* (aqueduct, chapel, fountain, convent, and church, respectively). They also include many architectural elements like *alpendre*, *janela*, *parede*, *porta*, *retabulo*, and *telhado* (porch, window, wall, door, altarpiece, and roof, respectively). Figure 4 depicts the most relevant articles for topics related to these terms. Notably, the significance of specifically Christian institutions is emphasized not only by some of these terms but also by others such as *irmandade*, *ordem*, and *padre* (brotherhood, order, and priest, respectively). Additionally, Figure 4a highlights an emphasis on buildings linked to economic activities, as indicated by the terms *engenho* and

fazenda (mill and farm, respectively).

Our topic analysis also indicates how the idea of cultural heritage privileged by the journal was closely connected to an organization of the Brazilian territory that derived from the colonial model implemented by Portugal. This is clear from Figure 5, which refers to topics containing terms associated with different units of territorial organization, such as *aldeia*, *cidade*, *estado*, *município*, *região*, and *vila* (village, city, state, municipality, region, and town, respectively), in Figure 5a and Figure 5b. Moreover, the behavior of topics related to terms connected to the idea of territorial defense, such as *forte* and *fortificação* (fort and fortification, respectively), is illustrated in Figure 5c and Figure 5d. Although the temporal trend seems to affirm the importance of the Brazilian territory, it is important to stress that the journal of the SPHAN/DPHAN has never given equal value to all Brazilian regions. On the contrary, only a few states located on the coast of the country are mentioned – usually through references to one or more of their most important cities –, namely *Bahia*, *Ceará (Fortaleza)*, *Minas Gerais (Mariana, Vila Rica – now Ouro Preto)*, *Pará (Belém)*, *Pernambuco*, and *Rio de Janeiro*.

4.4. Other Civilizational Matrices

An additional aspect of the topics indicated in Figure 2 deserves to be highlighted. Among these topics, one can note themes that, if not completely ignored in the critical reception of the journal, have not received due attention to date. Indeed, our analysis reveals, somewhat surprisingly, that articles published in the journal addressed themes that may be unexpected to several art historians. For instance, topics 2 and 4 demonstrate that the published articles also considered – albeit in a more circumscribed way – other civilizational matrices, such as the African (indicated by terms such as *escravo* and *negro* – slave and black, respectively – in topic 2), and specially the Indigenous ones (indicated by terms such as *indio* and *indígena* – indian and indigenous, respectively – in topics 2 and 4).

The STM analysis, as depicted in Figure 6, shows that both civilizational matrices are simultaneously predominant (due to topic 2) by at least 0.9765 in 2 articles, published in 1944 (CrtN) and 1956 (J.MR). Moreover, the analysis revealed that the Indigenous theme is exclusively predominant by at least 0.9512 in 5 articles, published in 1937 (HlAT), 1938 (RmnL), 1939 (CrLE), 1942 (GstC), and 1959 (SydV). Furthermore, other articles show significant influences from these themes (by at least 0.4472), with 1940 (AlbL) and 1944 (SrfL) showing influences due to topic 2, and 1937 (RMFdA), 1938 (EstP), and 1941 (GstC) showing influences due to topic 4.

In this sense, it is particularly noteworthy, for example, the presence in topic 4 of the term *pesca* (fishing) and, consequently, of the article by Raimundo Lopes entitled “*Pesquisa Etnológica sobre a pesca brasileira no Maranhão*” (Ethnological research on Brazilian fishing in Maranhão), published in the second issue of the journal. Lopes certainly pays attention to the materiality of the tools used by indigenous peoples of Maranhão in their fishing activities; but it is above all the activity of fishing itself that appears as the main theme in his article. This points to the fact that the journal of the SPHAN/DPHAN – at least in its early years – welcomed discussions that transcended the merely material character of the heritage analyzed. In Brazil, this type of concern with the also “immaterial” nature of heritage – with the *savoir-faires*, not only with the finished products – has only become really relevant since the 2000s [23].

These findings shed light on the more diverse than expected nature of the topics explored in

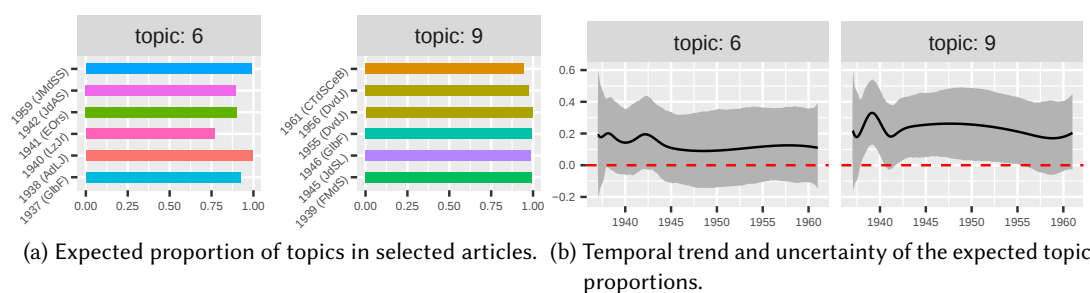


Figure 3: Profiling of two topics dealing with artistic production modalities.

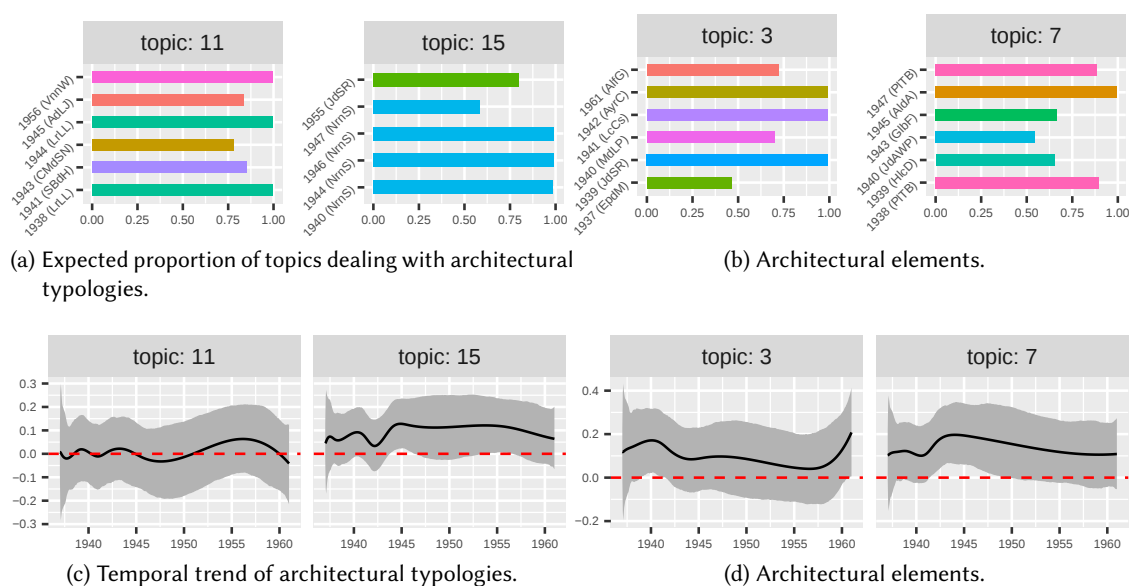


Figure 4: Profiling of topics dealing with architectural typologies and elements.

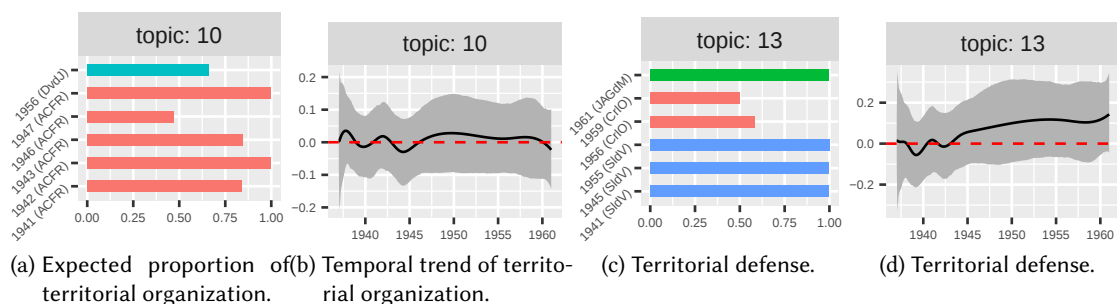


Figure 5: Profiling of topics dealing with space modalities.

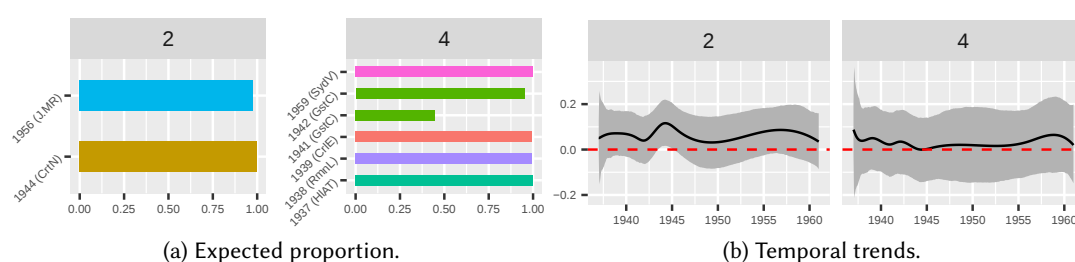


Figure 6: Profiling of topics dealing with other civilizational matrices.

the journal's articles, further enhancing its importance as a source for reflection on the diverse cultural matrices that have shaped Brazil's heritage and history.

5. Concluding Remarks

Through automated topic analysis of over 4700 pages of the journal of the SPHAN/DPHAN, we were able to unveil patterns and nuances of prevalent and less prevalent topics that provided valuable insights into the journal's content and thematic dimensions. First, our findings confirm the general perception that the journal indeed focused on the history of material civilization in Brazil during the Portuguese colonial period. Within this overarching theme, our automated analysis provided insights into the specific approaches taken by the articles, particularly highlighting architectural construction modalities and geographical locations of the heritage discussed therein. However, it is essential to acknowledge that the method proposed in this research required expert intervention to prepare the dataset for the automated analysis. Despite this limitation, the use of automated textual analysis proved to be a valuable tool in uncovering previously unnoticed aspects and challenging common assumptions about the journal's content.

Secondly, our analysis also stressed the underrepresentation of other civilizational matrices in the formation of the Brazilian nation, such as Indigenous and African ones. Our automated analysis uncovers another underrepresented theme related to distinct periods of Brazilian history beyond the colonial era. For example, topic 14 reveals a series of terms directly connected to the Academy of Fine Arts that was founded in Rio de Janeiro only at the beginning of the 19th century. This highlights that automated analysis, while not a replacement for in-depth qualitative investigations, can offer new possibilities for uncovering nuanced aspects and supporting subsequent qualitative examinations. By seamlessly integrating automated textual analysis and qualitative investigation, researchers can gain a more holistic understanding of Brazilian cultural identity and heritage preservation efforts. In this regard, future investigations could explore the utilization of large language models to enhance the automated analysis process.

References

- [1] A. Brey, Digital art history in 2021, *History Compass* 19 (2021) e12678. doi:10.1111/hic3.12678.
- [2] J. Drucker, Is there a “digital” art history?, *Visual Resources* 29 (2013) 5–13. doi:10.1080/01973762.2013.761106.
- [3] P. Fletcher, Reflections on digital art history, *caa.reviews* (2015). doi:10.3202/caa.reviews.2015.73.
- [4] R. A. de O. Lanari, O patrimônio por escrito: a política editorial do Serviço do Patrimônio Histórico e Artístico Nacional durante o Estado Novo (1937-1945), *Letramento*, 2010.
- [5] R. O. Ribeiro, Revista do Patrimônio Histórico e Artístico Nacional: a história da arte engajada na política de preservação no Brasil, Master’s thesis, Universidade Estadual de Campinas, 2013.
- [6] A. F. Silva, P. Faulhaber, Narrativas sobre o patrimônio: Rodrigo Melo Franco de Andrade, redes de sociabilidade e a escrita do patrimônio na revista do patrimônio (1937-1945), *Anais do Museu Histórico Nacional* 51 (2019) 150–173.
- [7] C. M. de C. Silva, Revista do Patrimônio: editor, autores e temas, Master’s thesis, Fundação Getúlio Vargas, Centro de Pesquisa e Documentação de História Contemporânea do Brasil, 2010.
- [8] L. S. Teixeira, Civilização material, história e preservação em afonso arinos, in: M. Chuva, A. G. R. Nogueira (Eds.), *Patrimônio cultural: políticas e perspectivas de preservação no Brasil*, Mauad X, Rio de Janeiro, 2012, pp. 47–57.
- [9] M. R. R. Chuva, Os arquitetos da memória: a construção do patrimônio histórico e artístico nacional no Brasil (anos 30 e 40), Ph.D. thesis, Universidade Federal Fluminense, 1998.
- [10] A. A. de M. FRANCO, *Desenvolvimento da civilização material no Brasil*, Ministério da Educação e Saúde, Rio de Janeiro, 1944.
- [11] O. Bonfait, A. Courtin, A. Klammt, Humanités numériques et histoire de l’art en france: état de la recherche et perspectives, *Histoire de l’Art* 87 (2021) 5–16.
- [12] T. L. Griffiths, M. Steyvers, Finding scientific topics, *Proceedings of the National Academy of Sciences* 101 (2004) 5228–5235.
- [13] J. Grimmer, B. M. Stewart, Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts, *Political Analysis* 21 (2013) 267–297. Publisher: Cambridge University Press.
- [14] M. Roberts, B. Stewart, D. Tingley, E. Airoldi, The structural topic model and applied social science, *Neural Information Processing Society* (2013).

- [15] A. Wasielewski, A. Dahlgren, Mining art history: Bulk converting nonstandard pdfs to text to determine the frequency of citations and key terms in humanities articles, in: S. Petersson (Ed.), *Digital Human Sciences: New Objects – New Approaches*, Stockholm University Press, Stockholm, 2021.
- [16] M. B. Rezende, B. Grieco, L. Teixeira, A. Thompson, Serviço do Patrimônio Histórico e Artístico Nacional (SPHAN) 1937–1946, in: *Dicionário IPHAN de Patrimônio Cultural*, IPHAN/DAF/Copedoc, 2015.
- [17] D. Williams, *Culture Wars in Brazil: The First Vargas Regime, 1930-1945*, Duke University Press, Durham & London, 2001.
- [18] Brasil, Lei n°. 378, de 13 de janeiro de 1937. dá nova organização ao ministério da educação e saúde pública, <https://www2.camara.leg.br/legin/fed/lei/1930-1939/lei-378-13-janeiro-1937-398059-publicacaooriginal-1-pl.html>, 1937.
- [19] R. M. F. d. Andrade, Programa, *Revista do Serviço do Patrimônio Histórico e Artístico Nacional* 1 (1937).
- [20] M. Steyvers, T. Griffiths, Probabilistic Topic Models, in: *Handbook of Latent Semantic Analysis*, Psychology Press, 2007.
- [21] D. M. Blei, Probabilistic topic models, *Communications of the ACM* 55 (2012) 77–84.
- [22] M. E. Roberts, B. M. Stewart, D. Tingley, stm: An R Package for Structural Topic Models, *Journal of Statistical Software* 91 (2019) 1–40.
- [23] M. C. L. Fonseca, Para além da "pedra e cal": por uma concepção ampla de patrimônio, in: R. Abreu, M. Chagas (Eds.), *Memória e patrimônio: Ensaios contemporâneos*, 2 ed., Lamparina, Rio de Janeiro, 2009, pp. 59–79.