

Historical Portrayal of Greek Tourism through Topic Modeling on International Newspapers

Eirini Karamouzi^{1,2}, Maria Pontiki^{3,4}, and Yannis Krasnikolakis⁵

¹University of Sheffield, ²American College of Greece,

³Panteion University of Social and Political Sciences, ⁴Athena Research Center, ⁵Grant Thornton

e.karamouzi@sheffield.ac.uk, m.pontiki@panteion.gr, yannis.krasnikolakis@gr.gt.com

Abstract

In this paper, we bridge computational linguistics with historical methods to explore the potential of topic modeling in historical newspapers. Our case study focuses on British and American newspapers published in the second half of the 20th century that debate issues of Greek tourism, but our method can be transposed to any diachronic data. We demonstrate that Non-negative Matrix Factorization (NMF) can generate interpretable topics within the historical period under examination providing a tangible example of how computational text analysis can assist historical research. Furthermore, we highlight the role of historical interpretation with regard to analyzing discourse dynamics through topic models. The results of our analysis provide interesting insights for academics and researchers in the field of (Digital) Humanities and Social Sciences, as well as for stakeholders in the tourism industry.

1 Introduction

Mass tourism has become a global phenomenon and a \$9 trillion industry which, in 2022, accounted for 7.6% of the world's GDP according to the World Travel and Tourism Council.¹ It is vital to the well-being of individuals and nations, and of cultures, economies and societies, proven by the absence of travel during the pandemic and its revival since. Tourism is now in question as climate change forces the world to reconsider air miles and ecological footprints. It is more

important than ever to understand tourism's history to inform decisions about its future. In this paper, we take Greece as a case study and we explore the country's transition to mass tourism from 1945 to 1989 for two reasons. First, it is widely accepted in tourism studies that modern international tourism began in earnest in the 1950s and grew exponentially thereafter, especially with the introduction of affordable package holidays (Zuelow, 2011). In Greece in 1950, visitor numbers were c. 33,000. Over the next four decades, that figure grew to c. 9 million. We also selected 1989 as an end date for practical research purposes since liberalization of air travel, end of Cold War divisions and the usage of the web changed the landscape of tourism and its promotion. There is also a methodological issue. In the second half of the 20th century, the newspapers played a central role in the political, economic and cultural life of Britain and the USA. For instance, the British read more newspapers per capita than any other people in the world (Bingham, 2010). Therefore analyzing the debates in these newspapers is an invaluable source of evidence for a wide range of historical investigations and in our case of Greek tourism.

Our focus is American and British newspapers, since the USA and UK were the main markets of tourists throughout this period, along with France and Germany. We experimented with LDA and NMF to uncover the main topics and narratives framing Greek tourism in international newspapers and explore the main textual representations of a tourist nation, their evolution over time, their cultural underpinnings, and how they were communicated to different audiences.

¹<https://wtcc.org/research/economic-impact>

The contribution of our work is two-fold; first, the extracted topics are evaluated both by a computational linguist and by a historian highlighting the crucial role of domain experts when interpreting topic modeling outputs. Second, the extracted topics are contextualized within the historical and political environment in which they appear, providing interesting insights about the historical representations of Greek tourism over the years, and about the development and the hallmarks of American and British tourism in Greece across different historical periods (from 1945 to 1989). The comparative analysis between the American and the British press reveals interesting insights including similar responses to specific events as well as notable differences between British and American tourism to Greece during the historical periods under examination. Overall, the results of our analysis can provide valuable information for academics and researchers in the field of (Digital) Humanities and Social Sciences, as well as for stakeholders in the tourism industry.

2 Background

Tourism has held a quasi-monopolistic status in Greece's development. While the promotion of tourism in Greece at times mirrored the agenda of the Greek state in the construction of its nation's image during its turbulent history of civil war, dictatorship, and democratization, it has also included the power of the market. After its emergence in the 1960s, almost 80% of Greek tourism has been inbound, which exemplifies the industry's exposure to international developments and events. How a nation depicts itself is not a value-free expression of identity, but a culmination of historical processes that 'reveal much about the social construction of space, cultural change, identity and discourse' (Pritchard & Morgan, 2001). The topic has been researched by visual artists, architects, ethnographers and economic historians of tourism who have looked at governmental archives, hotel/airline and tour operators repositories, posters, official and unsanctioned travel guides and personal travelogues and tourist documentaries and films (Tsartas, 2010; Vlachos, 2016; Alifragkis & Athanassiou, 2013). However, historical methods alone are insufficient for this study. By using

computational approaches, it can advance the field of study beyond anthropology, ethnography and sociology to understand the major themes of how Greece promoted itself to different audiences, at different times.

Furthermore, newspapers constitute valuable sources in historical research, providing windows into the past, but also posing challenges for historians to go through their sheer volume of information page-by-page (Yang, Torgtet, & Mihalcea, 2011). Recent advancements in the field of historical research have witnessed a growing interest in leveraging topic modeling techniques to capture historical trends (e.g. Newman & Block, 2006; Oiva, 2021) and study discourse dynamics diachronically in newspapers collections (e.g. Marjanen et al., 2021; Viola & Verheul, 2019). In the tourism field, topic modeling is usually applied to travelogues and user comments on travel websites (e.g. Pang et al., 2011) as an important tourist attraction profiling technique facilitating personalized attraction recommendation services (Huang et al., 2018). Other approaches focus on online review platforms and social media data to analyze consumer perceptions and (dis)satisfaction of visitors of hospitality and tourism-related products, organizations and services (e.g., Guo, Barnes, & Jia, 2017; Kim, Kim & Park, 2021). To the best of our knowledge, topic modeling on historical newspapers has not been employed before in tourism studies. In the field of Digital Humanities, digital methods have been used to study travel and networks², but not the construction, transfer or consumption of a country's tourist destination within its ideological, political and cultural context.

3 Datasets

3.1 Data Collection and preparation

Based on a set of specific keywords related to Greece and Greek tourism (Appendix C) we collected a total of 1099 news articles from 11 sources for the period 1945-1989, as presented in Table 1. The digitized newspapers were from ProQuest Historical Newspapers and GALE Primary Sources, while New York Times has its own archival repository with paid subscription that we utilised. In order to ensure a balance of

²See <https://grandtour.stanford.edu/>

viewpoints, our selection of newspaper sources includes a range of broadsheets, economic newspapers, left and right on the political spectrum and magazines.

	American	British
Newspaper (number of articles)	Chicago Tribune (81), Los Angeles Times (60), New York Times (461), Wall Street Journal (26), Washington Post (44), Vogue (23)	Financial Times (82), Guardian Observer (86), Telegraph (49), The Economist (41), The Times (146)
Total	695	404

Table 1: Number of articles per newspaper.

Then, we applied OCR to transform the articles into machine readable format. For the American press, we used the *Tesseract* tool. For the British press, the same tool could not be applied for most of the newspapers (i.e. The Times, The Financial Times, Telegraph, and The Economist), so we used *ocrmypdf*, a Python library that adds an OCR text layer to scanned PDF files, allowing them to be searched. To estimate the quality of the extracted texts, we used the *enchant.Dict("en_US")* dictionary available at the *enchant* library, to determine the percentage of words that belong to the English vocabulary. The algorithm calculated the ratio of the recognized English words to the total number of words in each article providing a quantitative measure of the OCR text quality, with the percentage for the American articles being 93.47%, and for the British, 92.22%. In the list of the non-vocabulary words, there were some misspelled words (e.g. ‘Speciglists’, instead of ‘Specialists’). To fix this issue we used the *pyspellchecker* Python library. After fixing misspellings, we observed that some of the remaining non-English words were names or places, so they should not be extracted from the text. We applied the Spacy Named Entity Recognizer to detect named entities using the ‘PERSON’ and ‘LOC’ (location) labels. The extracted terms were curated and cataloged within a non-English vocabulary list to ensure their retention within our corpus. Finally, we removed the list of non-vocabulary words left since it contained poorly extracted non-recognizable terms.

3.2 Data Organization

The data are organized in two main corpora per nationality (American and British), and further splitted according to specific periods. Breaking down news data into time windows that align with historically significant periods has previously been employed to uncover noteworthy historical trends (e.g. Yang, Torgtet, & Mihalcea, 2011; Hengchen, 2017). In our case, the following five periods have been set:

- A. 1945-1949: Reconstruction of the country and first signs of tourism.
- B. 1950-1966: Tourism takes off.
- C. 1967-1974: Dictatorship.
- D. 1975-1980: Democratisation & economic crisis.
- E. 1981-1989: PASOK government.

Period	American		British	
	articles	tokens	articles	tokens
1945-49	16	10683	22	80375
1950-59	267	156684	164	527021
1960-69	133	96507	103	261726
1970-79	110	116601	42	88751
1980-89	169	180825	73	129181

Table 2: Number of articles and tokens per period.

4 Topic Modeling

Topic models provide an effective way to draw insights from large-scale collections and are widely used in digital humanities and social sciences (Brauer & Fridlund, 2013; Marjanen et al., 2020) to uncover the most prevalent themes on different types of data ranging from books, newspapers, and academic journals to parliament proceedings and social media. Since the inception of the term ‘topic model’ along with the introduction of Latent Dirichlet Allocation (LDA) by Blei, Ng, and Jordan (2002), topic modeling research has evolved to address the challenges arising from different types of applications, since it became clear that not all algorithms are effective in all types of text (Churchill & Singh, 2022). In this paper, we experiment with LDA and Non-negative Matrix Factorization (NMF), two models that are very popular across various domains (e.g.

Pitichotchkophokhin et al., 2020; Egger & Yu, 2022).

The input for the topic models was preprocessed data. All the articles were lowercased, tokenized, part-of-speech tagged and lemmatized. Stop words were removed and bigrams were extracted to obtain topics with phrase-like keywords and not only single terms. After several iterations, the stop words list was enriched with more non content terms, and we decided to remove verbs and adverbs and to keep only nouns and adjectives. To find the best hyperparameters for LDA, we utilized Gensim's³ LDA Model, focusing on optimizing the number of topics and the alpha parameter. Alpha, representing the document-topic density, influences how many topics a document potentially has. We experimented with values like 0.1 (indicating low topic density per document), 'symmetric' (assuming an equal distribution of topics across documents), and 'asymmetric' (allowing for a varied distribution of topics). The number of topics was also varied to identify the optimal structure for our datasets. This approach led us to identify a total of 18 topics in the American articles ranging from 2 to 5 per period, and a total of 24 topics in the British articles ranging from 2 to 7 per period. The results were visualized using pyLDAvis⁴ to provide an intuitive understanding of the topics and their distribution.

For NMF, we employed the TfidfVectorizer from scikit-learn⁵ to construct the terms-documents matrix. This matrix represents the importance of terms in each document, with higher weights assigned to terms that are frequent in a specific document but rare across the entire corpus. To estimate the optimal number of topics, we computed the highest coherence scores for each dataset and sub-corpus using Gensim's CoherenceModel. This model evaluates the coherence of topics by measuring the degree of semantic similarity between high scoring words in the topic. We identified a total of 39 topics in the American articles ranging from 6 to 9 per period, and a total of 34 topics in the British articles ranging from 3 to 9 per period.

The human evaluators concluded that NMF generates more interpretable topics than LDA on the specific datasets. A possible reason could be the fact that LDA is a static approach, whilst NMF can capture topic evolution in temporal data. Another

possible explanation could be that, as reported in previous research (e.g. Young & Johnson, 2018), NMF performs better than LDA on a smaller number of documents. We plan to further explore the differences between the LDA and NMF outputs, considering also the unbalanced nature of our datasets (i.e. different number of articles from various newspapers for different periods). In the next section, we discuss findings based on the NMF results.

5 Results and Discussion

The topic modeling results for the American articles are presented in Appendix A and for the British articles in Appendix B. The extracted topics were qualitatively evaluated by a historian and a computational linguist. In this section we discuss the most important findings with the aim to highlight the potential of topic modeling in historical research, and the crucial role of domain experts when interpreting topic modeling outputs.

5.1 American Press

The results on the American press for the period **1945-1949** reveal 4 topics discussing the catastrophic damage wrought to the Greek economy during the Second World War, with tourism framed as potential means of partially mitigating Greece's balance of payments crisis, along with the restoration and cultivation of agricultural and industrial productivity (mainly of tobacco) (#A8). US aid provided to Greece and other European nations under the auspices of the Marshall Plan (Economic Recovery Programme) (#A1, #A6), following its inception in 1948 could in fact be used as a means of redeveloping a recipient's tourist industry despite the ongoing Civil War (1946-1949) (#A3). In topic #A7 which encompasses terms related to hospitality industry in Greece and other Mediterranean countries (e.g. hotel, restaurant, prices, currency), and in topic #A4 through terms reflecting aspects of transportation/travelings, one can detect the first signs of contemporaneous uptick in US travel to the Mediterranean with the majority of these journeys undertaken via cruise ships and other seagoing passenger vessels. Finally, topic #5 on Rhodes, recently returned to Greek sovereignty after Italian occupation, was given high priority as

³ <https://radimrehurek.com/gensim/>

⁴ <https://pypi.org/project/pyLDAvis/>

⁵ <https://scikit-learn.org/stable/>

an already established popular tourist destination with the necessary infrastructure provided by the Italian administration.

In the **1950-1966** period, we can identify the onset of the jet age and the shifts in US tourist market (#B2). By 1953, air travel had made the tourism business a 12-month affair, and facilitated long trips while fostering off-season travel through reduced fares. That was further fortified by cruise-related travel services (#B9). Americans were encouraged to vacation in Greece for reasons of affordability where they would be the ‘most for their money’ (#B3). This new reality pushed the Greek government in 1957 to put into operation a five year tourist investment plan, including the building of the Hilton hotel in Athens in addition to coastal areas near Athens (#B1, #B7). The focus on both the Delphi ancient site as well as the islands of Mykonos and Rhodes as tourist destinations (#B4), reflect how Greece was attempting to capture a larger portion of the growing international travel market by trumpeting the emergence of a ‘modern’ Greece that offered sea, sun and sand alongside its much-lauded antiquities and cultural politics. An interesting topic is #B6. Based on the top terms, the computational linguist recognised a topic discussing the legal status of Green-born citizens in the US, while the historian identified the issue of the Greek diaspora. Equally important are topics #B5, #B7, and #B8 that showcase the novel feature of the Greek tourism industry to target the diaspora in the USA and utilize its members for publicity events.

In the Dictatorship period (**1967-1974**), topics #C2, #C6 and #C4 indicate how the establishment of the regime of the Colonels in April 1967 did not hinder recently established relationships between US private investors and representatives of the Greek tourism and aviation sector. For example, a contract between the Greek government and Litton Industries Inc., pertaining to an ambitious 12-year ‘nation building’ project that focused *inter alia* on the development of tourism infrastructure within the western Peloponnese, was reaffirmed by the junta in short order. And while tourism to Greece to a ‘nose dive right after the coup’, it did not take long for normal service to be resumed with usual concerns of sightseeing, shopping and taverna comparings as we can notice to the topics discussing again the islands of Mykonos, Rhodes and this time also Kos as tourist

destinations in the Aegean Sea (#C1), on travel services (#C3), and heritage/gastronomy (#C7).

During the period of democratization and economic crisis (**1975-1980**) we identified 6 interpretable tourism related topics discussing this time Crete island as a tourist destination (#D1), hospitality services (#D2), cultural events (#D3), travel services (#D4), Athens and ancient Olympia (#D5) and touristic development of Pylos whilst for the first time we see environmental concerns over the touristification of places in Greece (#D6).

Finally, during the first socialist government of PASOK (**1981-1989**) along with the expected topics (#E1, #E5, #E4) discussing different aspects of Greek islands (i.e. sea, beach, town, village) and related travel (e.g. boat) and hospitality services (hotel, room, restaurant), we have two noteworthy developments. The depressive influence of a protracted economic recession in Europe, as indicated in topic #E6 on dollar and other countries continued to limit the growth of the Greek tourism industry until 1984. But the development of US tourism to Greece would, for the remainder of the decade, be defined primarily by debates concerning international terrorism, which achieved a degree of prominence hitherto unseen following the hijacking of TWA flight 847 shortly after its departure from Athens on 14 June 1985. The topics of cancellations (#E6) and hostage (#E8), refer to President Ronald Reagan’s issuance of an injunction to US tourists to boycott Greece until security measures against terrorist attacks at Athens airport were improved (#E2). The effects of this intervention were felt immediately. With travel agents in Athens reporting mass cancellations of American bookings (#E8), and tensions escalating between the socialist government of Andreas Papandreu and the Reagan White House. The Greeks called upon George Lois to launch a campaign and attract much needed international visitors and their tourist dollars (#E7, #E9).

5.2 British Press

For the period **1945-1949**, we identified 2 topics focusing on governance and politics (#A2), and the economy and banking sector (#A3). The coverage underlines how an understanding of the economic realities imparted a sense of pragmatism among war-weary Greeks, for whom the exigencies of getting their country’s economy back on its feet – which necessitated a

rejuvenation of the tourist trade – superseded partisan politics, even during a period of deep internal unrest. The British are more interested than the Americans in the Greek political realities and challenges of governing. However, the Marshall plan, as with the American press, takes center stage in relation to its potential to tourist development.

The period of the tourist boom of **1950-1966**, is linked to economic policies (#B5). The 1955 annual review of the Bank of Greece revealed that earnings from tourism had increased by 164% between 1952 and 1954 on account of the devaluation of the drachma by fifty percent in April 1953 (#B4). It was not just domestic economic policy, private enterprise from Germany for instance also began to feature prominently in the development of tourist infrastructure (#B4). Equally important was the role of cruises (#B7). The Greek government hoped to maximize the impact of a recently discovered medium through which the natural beauty of Greece could be transmitted to international audiences – Hollywood motion pictures, as discussed in #B6 topic. The topic of destinations (Athens and Mykonos Island), archaeological sites (Parthenon and Delos), as well as the mention of sea and feeling good, reflects the element of duality: the allusion to two different concepts of Greece; one characterized by the ancient ruins that attracted ‘ardent Hellenists’, another comprised of picturesque scenery and the promise of relaxation on the cheap and in the sun (#B3, #B1). While the exponential growth of the Greek tourist industry would depend to a significant degree on marketing the latter to swelling ranks of sun worshippers, it would appear that, for British audiences, the former remained the predominant conception of Greece, its appeal as a tourist destination predicated on the cultural and historical significance of its antiquities. The rest of the topics revolve around impact of regional political tensions and political upheaval caused by the outbreak of political violence in Cyprus in November 1955 and 1964 (#B4).

During the Dictatorship period (**1967-1974**), we can identify the impactful role of political events on tourism with the Turkish invasion of Cyprus in 1974 that brought about the fall of the junta featuring in #C4. However, during the junta and in 1969, some 1,609,000 visited Greece,

spending in excess of £80.6 million, making Greece (alongside Portugal) the fastest growing tourist center in Europe (The Economist, 1971). This influx, moreover, provided fresh impetus for renewed efforts to develop Greece’s tourist infrastructure (#C2), with particular emphasis placed on the building of new hotels, beaches, and camping ground, as well as communications (#C6).

During the democratization and economic crisis period (**1975-1980**) similar tactics were employed in prioritizing particular regions for the development of tourism (#D1) like Crete and Corfu (#D4) and the investment of vast sums of money to provide the necessary infrastructure. Tourist figures duly returned to expected levels in 1975 and continued to rise sharply thereafter, leading to overcrowding and overbooking in several tourist resorts (#D5). As with the American press, the British discuss how by 1979 Greece was approaching an alarming milestone – six million visitors, or ‘nearly two foreigners for every three Greeks’. The opening of the Porto Carras (#D6) tourist complex in Halkidiki in 1976, a conglomeration of hotels, villas, restaurants, swimming pools, golf courses and bathing beaches (#D4) represented a move towards a more sustainable approach to the development of the Greek tourism industry and a governmental effort to attract more ‘affluent’ holidaymakers and ‘yuppify’ (#D2). In addition, we can see an energy and oil industry related topic (#D3).

Finally, in the PASOK government period (**1981-1989**) we can see the British go back to the deeply rooted ideational associations with antiquity and cultural heritage (#E6, #E7). The economic downturn of the period is reflected on the topic of strikes in #E8 and the governmental response (#E1), as well as the impact on other European countries (#E3). We have again a general topic about islands and accommodation (#E2), but most importantly the issue of security and terrorism is discussed as with the American press (#E4).

5.3 Further Insights

In order to get additional insights from our datasets, we also calculated the term frequencies on the American and British articles for the whole period under examination (1945-1989). Overall, it is interesting that debates on Greek tourism don’t

prioritize the issue of cultural heritage, despite appearing in different topics in our analysis and giving us fresh insights on its meaning. As indicated by the following figures (font size indicates frequency), the American articles feature discourses on Greece as an opportunity for tourists to enjoy the country's 300 days of sunshine and the 1,416 islands available. Moreover, it confirms the analysis of the different topics that describe the American gaze on Greek tourism as a complex mosaic, ranging from fascination with antiquities to the natural beauty of landscapes and the warmth and hospitality of the people, depending on the needs of the audience across time.

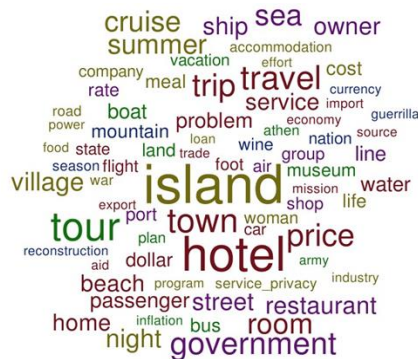


Figure 1. Summary of the most frequent terms in the American dataset.

The British articles diachronically focus on the Greek economic scene and the debates are quite consistent around governmental policy. There is preoccupation with Greece's investment in infrastructure, both public and private, with particular emphasis placed on the building of new hotels, roads and beaches. Also great attention is paid to the economic impact of the tourist industry and its dependency on international developments.



Figure 2. Summary of the most frequent terms in the British dataset.

Both American and British newspapers clearly show how the cultivation of Greece's tourist industry was seen as an economic imperative and an effective weapon in the battle against the country's perennial trade deficit and balancing the books. This was an orchestrated state endeavor as reflected in several topics in both USA and UK that focus on state-induced developments of a nation-wide network of modern leisure facilities, including hotels, motels, tourist pavilions, organized beaches and other infrastructural projects (roads, port facilities). The British are more interested in the Greek political and economic scene than the Americans in discussing tourism. Although technological advancements such as the onset of the jet age are seen as turning points on the increase of the tourists flows to Greece, it's interesting to see the complete omission of the role of other non-governmental stakeholders such as Olympic Airways, the country's flag-carrier from 1957 to 2009, that have been highlighted in most studies of the history of Greek tourism.

Beyond the economic function of tourism, the computational investigation in the foreign press confirms tourism's cultural, ideational and societal features. The results showed that most topics were on cultural politics of tourism promotion that should further encourage scholarship on tourist policy beyond the premises of the industry's significant multiplier effect on economic activity, which has been the focus of most of tourism literature.

The quantitative analysis of the newspapers showed how the debate on Greece as a tourist destination contributed to different notions of Greekness going back to the inception of the Greek state that sought to establish continuity between antiquity and modern Greek culture (Beaton, 2021). Early on, the Greek state utilized antiquity and the classical past as the cornerstone of the country's tourism policy, evident in several topics. However, our mixed methods research problematizes this cultural dualism following recent trends in Greek historiography that encourage more attention to diversity and plurality in the cultural politics of the country (Tziouvas, 2021). Moreover, our research shows that the Greek tourism product was far from monolithic and was heavily dependent on international developments and changing attitudes of the tourist audiences.

The development of British tourism to Greece during 1945-1989 was characterized by a major shift from cultural tourism towards low-budget, mass tourism of the ‘sun, sea and sand’ variety. This transition was, for the most part, in keeping with the imperatives of the Greek tourism industry, which recognised that such a reorientation was required in order to sustain growth. However, the transformation (or transmogrification) of British tourism to Greece was so rapid and far reaching that, by the mid-1980s, there was a revival of the cultural tourism that had predominated during the post-war years. The growth of American tourism vis-à-vis Greece during the same period was a somewhat messier, amorphous process. Themes such a fascination with antiquities, the natural beauty of landscapes, and the warmth and hospitality of the Greek people were (inter)changeable depending on the perceived preferences of US audiences at the time. Common flashpoints, such as the military coup that installed the Regime of the Colonels and the spate of terrorist incidents during the 1980s, also yielded similar responses among British and American audiences, although as regards the fallout from terrorist violence, Greek stakeholders were undoubtedly more proactive in their attempts to assuage the concerns of US tourists through a series of publicity-related measures.

In fact, it could be argued that a major difference between the development of US tourism to Greece and the contemporaneous growth of the British variant was the more proactive efforts of representatives of the Greek tourism industry within the US to proselytise on behalf of their country. Another notable difference between British and American tourism to Greece during the period in question was the ‘ethnic’ tourism of the Greek diaspora in the US, for which no equivalent could be readily identified within Britain.

6 Conclusions

In this paper, we bridge computational linguistics with history with the aim to explore the potential of topic modeling in discovering significant topics and historical trends in newspapers on the representations of Greek tourism during the period 1945-1989. Overall, the qualitative evaluation of the extracted topics revealed that most of the NMF topics are interpretable. The computational linguist was able to provide

descriptive labels (e.g. economy, transportation) to the majority of the topics, while the historian was able to identify more topics that required domain knowledge and also to contextualize them with the historical period under examination. Therefore, and in accord with Yang, Torgtet, & Mihalcea (2011), we agree with Block (2006) that “topic simulation is only a tool” and it is essential that an expert in the field contextualizes these topics and evaluates them for relevancy.

Furthermore, by contextualizing the extracted topics within the historical and political environment in which they appear, we highlighted the role of historical interpretation with regard to analyzing discourse dynamics through topic models. The analysis of our findings provides interesting insights about the main textual representations of Greece as a tourist nation, their evolution across different historical periods, their cultural underpinnings, and how they were communicated to different audiences, and offer a tangible example of how computational text analysis can assist historical research.

7. Limitations

The findings discussed in this paper are based on the NMF results on our datasets. We will further explore the capacity and the differences of both methods (LDA and NMF) also considering the size and the unbalanced nature of our datasets (i.e. different number of articles from various newspapers for different periods). A further limitation of our datasets arises from the OCR process; in addition to the quantitative evaluation of the OCR extracted texts, a qualitative one is also needed to check for irrelevant texts that might be included in our datasets due to unclear separation between different articles on the same page, thus affecting the quality of the topic modeling results (i.e. generating irrelevant or non-interpretable topics). In addition, the data collection method introduces limitations to our datasets, since the selected keywords may not capture the complexity of tourism as a multifaceted concept, other relevant articles might have been written using different terminology, and, given that we are dealing with historical data, some keywords may become outdated as language evolves and/or new terms may emerge. Taking into account the above limitations, the results presented in this paper are based on the specific articles that have been retrieved and analyzed from each newspaper; our

findings cannot be considered representative of the coverage and the editorial stances of each newspaper included in our analysis for the period under examination and cannot be generalized to the whole spectrum of the American and the British press.

Acknowledgments

The authors are grateful to G.Kakavas, A. Vavadakis, C.Fragkiadiakis and A.Tsakiri for their valuable contribution in the data processing tasks. The work presented in this paper is supported by UKRI, Research England.

References

- Stavros Alifragkis, and Emilia Athanassiou. 2011. Educating Greece in Modernity: Post-War Tourism and Western Politics. *The Journal of Architecture*. 18. 10.1080/13602365.2013.838285.
- Roderick Beaton. 2021. *The Greeks: A Global History*. Basic Books, London, ISBN10: 1541618297.
- Adrian Bingham. (2010). The Digitization of Newspaper Archives: Opportunities and Challenges for Historians. *Twentieth Century British History*, 21(2), pages 225–231.
- David M. Blei, Andrew W. Ng, and Andrew M. Jordan. 2002. Latent Dirichlet Allocation. In *Advances in Neural Information Processing Systems (NIPS)*, T. G. Dietterich, S. Becker, and Z. Ghahramani (Eds.). 601–608.
- Sharon Block. 2006. Doing More with Digitization: An Introduction to Topic Modeling of Early American Sources. *Common-Place*, 6(2), January.
- Rene Brauer and Mats Fridlund. 2013. Historicizing topic models, a distant reading of topic modeling texts within historical studies. In *International Conference on Cultural Research in the context of Digital Humanities*, St. Petersburg: Russian State Herzen University.
- Rob Churchill and Lisa Singh. 2022. The Evolution of Topic Modeling. *ACM Computing Surveys*. 54. 10.1145/3507900.
- Roman Egger and Joanne Yu. 2022. A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts. *Frontiers in Sociology*. 7. 10.3389/fsoc.2022.886498.
- Yue Guo, Stuart S. Barnes, and Qiong Jia. 2017. Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tour. Manag.* 2017, 59, 467–483.
- Simon Hengchen. 2017. When Does it Mean? Detecting Semantic Change in Historical Texts. Ph.D. thesis, Universite libre de Bruxelles.
- Chao Huang, Qing Wang, Dongui Yanga, and Feifei Xu. 2018. Topic mining of tourist attractions based on a seasonal context aware LDA model. *Intelligent Data Analysis*. 22. 383-405. 10.3233/IDA-173364.
- Woohyuk Kim, Sung-Bum Kim, and Eunhye Park. 2021. Mapping Tourists' Destination (Dis)Satisfaction Attributes with User-Generated Content. *Sustainability*. 3(22):12650.
- Jani Marjaneen, Elaine Zosa, Simon Hengchen, Lidia Pivovarova, and Mikko Tolonen. 2020. Topic Modelling Discourse Dynamics in Historical Newspapers. *ArXiv, abs/2011.10428*.
- David J. Newman and Sharon Block (2006). Probabilistic topic decomposition of an eighteenth-century american newspaper. *Journal of the American Society for Information Science and Technology*, 57(6), 753–767.
- Mila Oiva. 2021. Topic Modeling Russian History. In: Gritsenko, D., Wijermars, M., Kopotev, M. (eds) *The Palgrave Handbook of Digital Russia Studies*. Palgrave Macmillan, Cham.
- Yanwei Pang, Qiang Hao, Yuan Yuan, Tanji Hu, Rui Cai, and Lei Zhang. 2011. Summarizing tourist destinations by mining user generated travelogues and photos. *Computer Vision and Image Understanding* 115 (3), pages 352 – 363, special issue on Feature-Oriented Image and Video Computing for Extracting Contexts and Semantics.
- Pimpitcha Pitichotchkokphokhin, Piyawat Chuangkrud, Kongkan Kalakan, Boontawee Suntisrivaraporn, Teerapong Leelanupab, and Nont Kanungsukkasem. 2020. Discover Underlying Topics in Thai News Articles: A Comparative Study of Probabilistic and Matrix Factorization Approaches. 759-762. 10.1109/ECTI-CON49241.2020.9158065.
- Annette Pritchard, and Nigel Morgan. 2001. Culture, identity and tourism representation: Marketing Cymru or Wales?. *Tourism Management*. 22, pages 167-179. 10.1016/S0261-5177(00)00047-9.
- Dimitris Tziouvas. 2021. *Greece from Junta to Crisis: Modernization, Transition and Diversity*. Bloomsbury Publishing, London.
- Paris Tsartas. 2010. *Greek Tourism Development* [in Greek] (Athens, 2010).
- Lorella Viola and Jaap Verheul. 2019. Mining ethnicity: Discourse-driven topic modelling of immigrant discourses in the USA, 1898–1920. *Digital Scholarship in the Humanities*, Volume 35 (4), pages 921–943.

Angelos Vlachos. 2016. *Tourism, and the public policies in Contemporary Greece, 1914-1950* [in Greek] (Athens, 2016).

Tze-I Yang, Andrew Torget, and Rada Mihalcea. 2011. Topic Modeling on Historical Newspapers. In *Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, pages 96–104, Portland, OR, USA. Association for Computational Linguistics.

May Young, and David Johnson. 2018. A Comparison of LDA and NMF for Topic Modeling on Literary Themes. UBC Wiki: https://wiki.ubc.ca/Course:CPSC522/A_Comparison_of_LDA_and_NMF_for_Topic_Modeling_on_Literary_Themes

Eric E. G. Zuelow. 2011. *Touring Beyond the Nation: A Transnational Approach to European Tourism History*. Routledge, London.

Appendix A. Topic modeling results on the American articles.

topic ID	Top terms
A. Reconstruction of the country and first signs of tourism (1945-1949)	
#A1	war marshall_plan economic balance aid import state export price united foreign
#A2	like text document street letter old state despite mountain dress night little right
#A3	guerrilla five marshall aid program ration year army plan need current vehicle
#A4	service passenger air athens day new_york france army office cargo piraeus national
#A5	island italian currency problem beauty bank italy state occupation cause considerable point recovery
#A6	nation food loan united national economy industrial international special president resource truman
#A7	europe food business hotel france little situation restaurant italy price currency men problem
#A8	tobacco olive worth product germany export dollar government oil price united central living
B. Tourism takes off (1950-1966)	
#B1	hotel athens site festival ancient accommodation room price government road cost rate delphi
#B2	line passenger air ship atlantic flight charter vessel airline state united service
#B3	million foreign export income import common market economy economic trade industry government

#B4	island town rhodes ancient sea water cruise church mykonos aegean ruin small old
#B5	owner reproduction historical newspaper request chicago mayor tribune million dollar
#B6	draft united state regulation citizen greek born embassy citizenship law
#B7	publication condo document request white hilton database new_york publisher text woman
#B8	timesmachine home subject company reprint store date service privacy
#B9	travel new_york tour trip cruise agent round fare hotel jet
C. Dictatorship (1967-1974)	
#C1	island boat mykonos sea harbor rhodes village beach aegean ko water small town
#C2	junta military political government colonel regime communist athens coup new politics press state
#C3	cruise tour travel new_york ship agent line timesmachine service air travel agent
#C4	litton_contract development government investment nation project international development litton_industry proposal president industry
#C5	event reproduction owner reproduction historical newspaper historical owner million national newspaper period increase request historical request
#C6	million onassis refinery government nacho export loan oil foreign economic investment economy growth
#C7	acropolis athens museum street food ancient foot square like restaurant marble parthenon ate
D. Democratisation and economic crisis (1975-1980)	
#D1	island town village villa beach crete boat small street harbor water old hotel
#D2	departure transfer trip jet baggage round daily escort tax hotel continental breakfast
#D3	festival opera theater ballet ticket music orchestra athens concert program subject performance soloist
#D4	travel tour new_york fare price office cost vacation million charter europe
#D5	ancient bus temple ruin acropolis game athens museum site olympia mountain modern roman
#D6	environment development cultural touristic social economy pylos physical pollution economic natural environmental dilemma
E. PASOK government (1981-1989)	
#E1	island town beach village hotel small like church room sea restaurant boat old
#E2	security airport travel terrorist terrorism flight cancellation passenger airline athens twa incident president

#E3	timesmachine cruise home reprint store service privacy company tour brochure search archive caribbean
#E4	museum acropolis art street collection plaza athens admission open hall square folk restaurant
#E5	version section print archive parthenon turtle edition headline national
#E6	dollar europe travel european hotel france britain million price terrorism tour italy paris
#E7	document request condo film database text publisher publication image movie caption
#E8	philharmonic orchestra government hostage official sour state basis minister concert united cancel
#E9	lois advertising agency campaign commercial celebrity account unusual broadcast thing print born

Appendix B. Topic modeling results on the British articles.

topic ID	Top terms
A. Reconstruction of the country and first signs of tourism (1945-1949)	
#A1	government new service national policy night nation state lord office communist minister london
#A2	price profit tax million stock div gold steel market ordinary net dividend bank
#A3	guardian observer rhodes owner reproduction request historical newspaper village italian manchester
B. Tourism takes off (1950-1966)	
#B1	temp rain wind moderate shower cloudy editor government university deg sunny work midday
#B2	oil canadian stock steel company correspondent price industry canada market bank montreal sale
#B3	hotel island athens guardian mykonos classical class sea parthenon delos room cheap good
#B4	government minister correspondent state united party prime president soviet german police general
#B5	profit bank company dividend market increase account rate investment ordinary balance tax stock
#B6	news london daughter church engagement film work music service son daily salary car
#B7	cruise guinea travel agent hellenic southampton ship palma passenger madeira voyage london
#B8	allowance guardian observer owner reproduction newspaper travel historical sterling

#B9	hide cassock son spoof bros ord brit store stamp duty motor metal bargain
C. Dictatorship (1967-1974)	
#C1	government party minister new police vietnam state president student united court soviet press
#C2	million increase credit rate economy growth investment economic bank development deposit earnings
#C3	hotel cruise brochure travel beach island sun car agent street london villa sea
#C4	turkish cyprus cypriot base force united nation island deceit troop minister military
#C5	guardian newspaper observer historical reproduction owner hotel request
#C6	hotel development bed automatic astir athens ote island telecommunication beach national private five year
#C7	company industry market price new business profit investment plant group chemical government gold
D. Democratisation and economic crisis (1975-1980)	
#D1	hotel beach observer guardian like town island historical newspaper reproduction shop request night
#D2	party union government minister state french soviet eec leader foreign gas general france
#D3	bank ship hotel fleet foreign energy investment shipping oil million industry banking owner
#D4	corfu flower village club tour yacht mount golf near hotel available bird crete
#D5	london tel book street car brochure mon mile blue air travel great
#D6	arras porto village development island project service settlement hotel mainland cohen adrienne
E. PASOK government (1981-1989)	
#E1	minister government soviet president prime new state force party foreign prime minister talk nudist
#E2	island beach hotel town ferry road guide bay guardian village room observer like
#E3	rate germany jan australia france canada japan bank belgium billion holland sweden index
#E4	terrorist bomb security guardian hijack attack terrorism ship twa athens agency sentence group
#E5	hotel brochure tel travel london night flight tour company sale car football free
#E6	site cruise athens island crete ancient archaeological museum aegean eec professor problem foreign
#E7	rhodes stone fist mercury archaeologist limestone smith antiquity nationality statue bronze god kill

#E8	union government party worker strike company communist socialist solidarity social pay new employer
#E9	turtle austrian bay banana strauss egg chess bavaria breed steer west austria gentleman

Appendix C. Keywords used for Data Collection.

Greece Tourism
Greek Tourism
EOT/ GNTO
Greece-Olympic Airways
Greece-Cruises
Greece- Tourism exhibition
Greece- sun sea sand food
Greece-Mediterranean
Greece diaspora-tourism