The concept of nation in nineteenth-century Greek prose fiction through computational literary analysis

Fotini Koidaki
Department of Modern Greek
and Comparative Studies
Aristotle University of Thessaloniki
54124 Thessaloniki – Greece
coidacis@gmail.com

Despina Christou
Department of Informatics
Aristotle University of Thessaloniki
54124 Thessaloniki – Greece
christoud@csd.auth.gr

Aikateriki Tiktopoulou
Department of Modern Greek
and Comparative Studies
Aristotle University of Thessaloniki
54124 Thessaloniki – Greece
atiktopo@lit.auth.gr

Grigorios Tsoumakas
Department of Informatics
Aristotle University of Thessaloniki
54124 Thessaloniki – Greece
greg@csd.auth.gr

Abstract

How can the construction of national consciousness be captured in the literary production of a whole century? What can the macro-analysis of the 19th-century prose fiction reveal about the formation of the concept of the nation-state of Greece? How could the concept of nationality be detected in literary writing and then interpreted? These are the questions addressed by the research that is conducted in this paper and which focuses on exploring how the concept of the nation is figured and shaped in 19th-century Greek prose fiction. We propose a methodological approach that combines well-known text mining techniques with computational close reading methods in order to retrieve the nation related passages and to analyse them linguistically and semantically. The main objective of the paper at hand is to map the frequency and the phraseology of the nation-related references, as well as to explore phrase patterns in relation to topic modeling results.

1 Introduction

Even though Literature is not the accurate representation of reality, the literary text is actively participating in the formation of the cultural identity of the community in which it is written and read. As far as the national identity is concerned, the literary discourse is interwoven with the processes of formation and negotiation of the nationalistic ties of a community, since it not only is embodied through the most important nation-building tool, Language, but reflects and, at the same time, shapes the collective conscious. The paper at hand presents and discusses the data-set, as well as the results, of a research project that was focused on exploring and analysing how the concept of nation is figured in Greek prose fiction of the 19th century.

The 19th century is of considerable cultural and political importance for Greece since this was a decisive period for the flourishing of Greek letters and for the course of Greek history in general. The 19th century was the century of Greek ethnogenesis, which had a great influence on the European movement of nation-states (Beaton, 2021). The Greek revolution of 1821 was the first successful liberal-national uprising in Europe because, after the fall of Napoleon and the Congress of Vienna, the authoritarian monarchies had been restored in the continent. The Greek War of Independence that launched exactly 200 years ago, ended the 400 years of Ottoman rule in the Greek territory, and resulted in the establishment of the first Hellenic Republic that led to the formation of the modern state of Greece. During the 19th century and in parallel with the building of independent Greece, typography emerged and flourished in the newly-born state. Newspapers, journals, and publishing houses appeared and multiplied over the century, revealing this way the escalation of publishing activity in Greece and the consequent flowering of

75

letters (Pouliasis, 2021). The explosion of literary publishing, which scaled up especially around 1880-1900, as well as the increase printing industry is evidence of the rise of literacy levels among Greek citizens. At the same time, the role of literary people in the Greek Language Controversy that arose before the Revolution, the refuge that several revolutionists found in literature and particularly in self-narration, as well as the connection established between national and literary language, provide evidence demonstrating the service of literature to the national purpose.

In this context, our research aims to explore how the nation concept is figured in the 19th-century Greek prose fiction in order to address questions regarding the presence of the concept of nation in novelistic tradition, but also the discussion and the conceptualisation of it. The rest of this paper is structured as follows. Section 2 presents background material concerning computational analysis of the notion of nation in literature. Next, we outline the dataset generated (Section 3), the methodology developed (Section 4), as well as the findings of our research (Section 5). Finally, Section 6 presents the conclusions of our work.

2 Background

The 19th century is the century of the formation of nation states. The relation between literature and (romantic) nationalism has been discussed for almost a century (Bradsher, 1921; Giffin, 1945; Fisher, 1980). Literature is known to be one of the major contributory factors in the nation-building process, both because it is embodied through Language -that is one of the most important characteristics of collective identity- and because it has the power of building a national audience. In the case of the creation of the Greek nation state, researchers from different fields have focused on aspects of the national character's building in the pre-revolutionary and post-revolutionary years and have commented on the reflection of these aspects in the periodical press of the 19th century, in texts of various types written by Greek and European scholars during this period, in specific authors and in selectively chosen literary texts.[1] The investiga-

tion of the issues of nation-building and national self-determination in a corpus of modern Greek prose of the 19th century is attempted for the first time here.

On the other hand, the attempts to computationally analyse the representation of nationality in literature are limited. For example, Weenink (2018) explores the national trends in Gothic fiction using topic modeling along with a qualitative approach of close reading and contextualization, while Erlin et al. (2021) count the geographic mentions (using NER models, part-of-speech tagging, place names gazetteer) and national phrases (taken form the "history of" Wikipedia page) in narration to find out the level of national attention in contemporary European fiction across multiple national contexts. To our knowledge, we are the first to computationally analyse the concept of nation in modern Greek fiction of the 19th century.

3 Data

To explore the concept of the nation in 19th century Greek prose fiction, we needed to create a literary corpus because, as yet, there are no corpora of 19th century Greek prose fiction available. Considering the project timescales and the financial factors, we built a micro-corpus, which we digitised and further processed to develop a final dataset that could provide sufficient knowledge to answer the questions posed by this project.

Among the corpus-building criteria that we defined, and which were related to the language, the date, and the type of publication, the comprehensive and balanced representation of the 19th-century Greek novelists was very important. Therefore, we didn't limit our corpus to the literary canon. Instead, we decided to include every novelist published in the Greek language during the 19th century, such as lesser-known artists, novelists publishing in districts outside the capital of Athens, a foreign author who wrote in the Greek language and printed outside Greece, as well as an unknown author. However, because the literary production of each author differs in size, we decided to select one title of each novelist to save time and maintain balance. Consequently, we have chosen the most representative work of each author, while taking into account the publication dates to ensure enough samples from every decade of the century. Finally,

---

[1]From the rich bibliography, we mention here the volume edited by Roderick Beaton and David Ricks Beaton and Ricks (2009), the monographs by D. Tziovas Tziovas (2017) and A. Politis Politis (2017) and the two volumes of the conference proceedings entitled Greekness and Diversity? Cultural Mediation and "National Character" in Nineteenth Century,

edited by A. Tampaki, and Ou. Polikandrioti Tampaki and Polikandrioti (2016).

we built-up a catalog of 91 novels, corresponding to 90 novelists and an anonymous title. The 91 novels are all written in the Greek language and published inside and outside Greece or Greek-speaking territory between 1811 and 1901.

The final corpus, which corresponds to more than 16000 pages, was digitised on the Transkribus[2] platform with an OCR model that we trained by transcribing a 150-page (35.223 words) sample of the three most common fonts of our corpus (Didot, Porson, Teubner). Table 1 shows our OCR model character errors.

Metric	Train	Validation
CER	0.62%	1.44%

Table 1: OCR's Character Error Rate (CER) on train and validation sets

To reduce the noise that would affect the accuracy of the text mining tasks, we massively processed the OCRed files by treating cases such as hyphenation, watermarks, repetitive page headers, additional noisy characters, etc. Figure 1 demonstrates word-counts distribution in the 19th century after the post-processing phase.
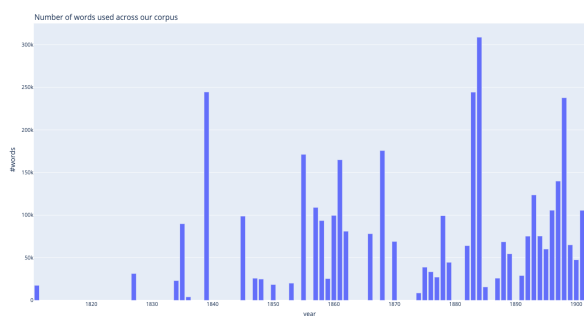

Figure 1: Word counts distribution in the 19th century

As we focus on such a specific topic as the nation is, we decided to prioritise a keyword-based passage retrieval approach, since this would allow us to collect and analyse all the nation-related passages contained in our corpus. The main idea was to establish a knowledge base featuring all the direct references to the concept of the nation comprised in Greek prose fiction, and which could be further studied, enriched, and analysed to answer the questions our project addresses. Therefore, the first step was to design the keyword query by defining the keywords of interest. Considering the available Greek dictionaries of the 19th century, we

defined seven keyword-stems related either to the nation in general or to the Greek nation in particular (translated from Greek language: nation, homeland, tribe, genus, Hellas, Greek, Romios [3]). For each of these keywords, we built a vocabulary, where we specified all the different grammatical forms derived from the seven stems, their possible spelling variations, as well as the unacceptable lexical types that may also derive from the same stems. These vocabularies allowed us to capture as many tokens as possible regardless of those including typos, spelling mistakes, idiom variations, or OCR noise.

The retrieved passages were organised on a tabular dataset and further enriched with various meta-information related to the passage-itself, the related keyword, and its parent document. [4] The passage retrieval was defined at the sentence level and not at the paragraph or phrase level, as the sentence, which is the smallest grammatical unit expressing a complete thought, would provide us with a sufficient sample to observe the concept of the nation in its context. Therefore, for each keyword captured, the query returned the complete sentence containing it alongside the matched keyword-stem and other useful meta-information. This suggests that in cases where more than one keyword was located in the same sentence, as many entries were created as the keywords captured in the sentence. Therefore, to monitor the findings' density and distribution in the documents, we assigned numeric identifiers to the sentences ("sentence-id") to observe the references through sentences. In the first step, a typical entry comprised the matched keyword, the complete grammatical sentence containing it, the "sentence-id", as well as useful metadata about the parent document (title, author, publication date, total sentence count, total words count).

However, the dataset was evaluated and further enriched. We repeatedly evaluated at least a random 10 percent of every keywords' results in order to fine-tune and optimise the keyword-vocabularies of the query. Based on the deficiencies observed, we decided to further evaluate the findings of two specific keywords (tribe, genus) in order to reduce noise caused by the findings that were not related to the concept of the nation. Having denoised the results as much as possible, we semantically enriched further the dataset with the geographical entities

[3] The Orthodox Christian Greek-speaking citizens of Ottoman empire

[4] Available upon acceptance.

that we automatically extracted from the retrieved passages. The final, evaluated and enriched version of the dataset is comprised of 8015 entries capturing and annotating individual references to the concept of the nation in the corpus of 19th-century Greek prose fiction.

4 Methodology

The generated dataset forms a topic-specific knowledge base that brings together all the nation-related references captured in the ninety-one nineteenth-century Greek novels presented above. To address the questions posed by the research, we subjected the dataset both to a statistical and linguistic analysis by combining well-known text mining techniques.

For instance, by grouping and filtering the dataset based on the collected metadata, we managed to develop a successively better picture about the frequency of the nation-concept occurrence and its presence in the corpus of the 19th century in general.

In the following, we deepened into a semantic analysis of the nation-related references to further understand the context describing the concept of nation. Precisely, we extracted the collocations around each nation-related keyword and all keywords to better understand the discussed topics in each keyword and their entire set, respectively. The lack of language processing tools (pos tags, stopwords) for the specific Greek language forms (Dimotiki and Katharevousa) of the 19th century - also characterizing our corpus - led us to combine tools developed for ancient- and modern- Greek language. For instance, the lack of pos taggers for the specific era guided us to enhance the stopword list by combining ancient- and modern-Greek stopwords lists, which we also enhanced with common OCR misspellings.

At last, we extracted clusters of words, i.e., words that frequently occur together, from the passages using topic modeling. Precisely, we used the Latent Dirichlet Allocation[5] (LDA) (Blei et al., 2003), the prevailing generative probabilistic topic model, to detect the thematic information of our archive. After hyper-parameters tuning, we concluded in fourteen (14) topics, among which there exists a topic that seems to be highly related to the concept of the nation-state. Finally, in order to read

---

[5]We used the Gensim (Rehurek and Sojka, 2011) python library

and further deepen our analysis on the topic modeling results, we exploited the knowledge gained from all the above-mentioned data-mining techniques.

5 Analysis

5.1 The presence of the concept of nation in the corpus

Dataset sorting, grouping and filtering provided interesting insights into the concept of nation in the 19th century Greek prose fiction corpus. The aggregation of data based on variables such as "keyword", "date" or "author" provided different perspectives on the frequencies of the nation-related references retrieved from 19th-century Greek fiction.

In particular, when we attempted to measure the occurrence of the concept of the nation in each document, we noticed that the keywords' occurrences were not always directly proportional to the length of each book. In fact, we compared the total number of sentences containing nation-related references (see in Figure 2 the red bars with y axis to the right) to the total length of each document measured in sentences (see in Figure 2 the cyan bars with y axis to the left). A sentence, of course, may have contained more than one reference to the nation. However, each sentence containing nation-related references was conceived as a single instance of a complete reference to the concept of the nation. We used the sentence as a measurement unit to create measurable quantities and properly map the complete references to the concept of the nation in our corpus. For example, Figure 2 shows that some shorter documents contain more references than longer ones, thus suggesting that the document's length should also be taken into account in our effort to quantify the presence of the concept of nation in the corpus.

The "sentence-id" was the variable that allowed us to measure our findings at the sentence level. Using the aforementioned measurements we estimated the rate of nation-related sentences for each document to observe the extent to which each novel refers to the concept or even the topic of the nation. However, when we generated a timeline graph of the reference rates (see Figure 4) we got a completely different view compared to that of the total number of the tokens retrieved (see Figure 3). More precisely Figure 3 presents the total number of tokens per publication date, while Figure 4 shows
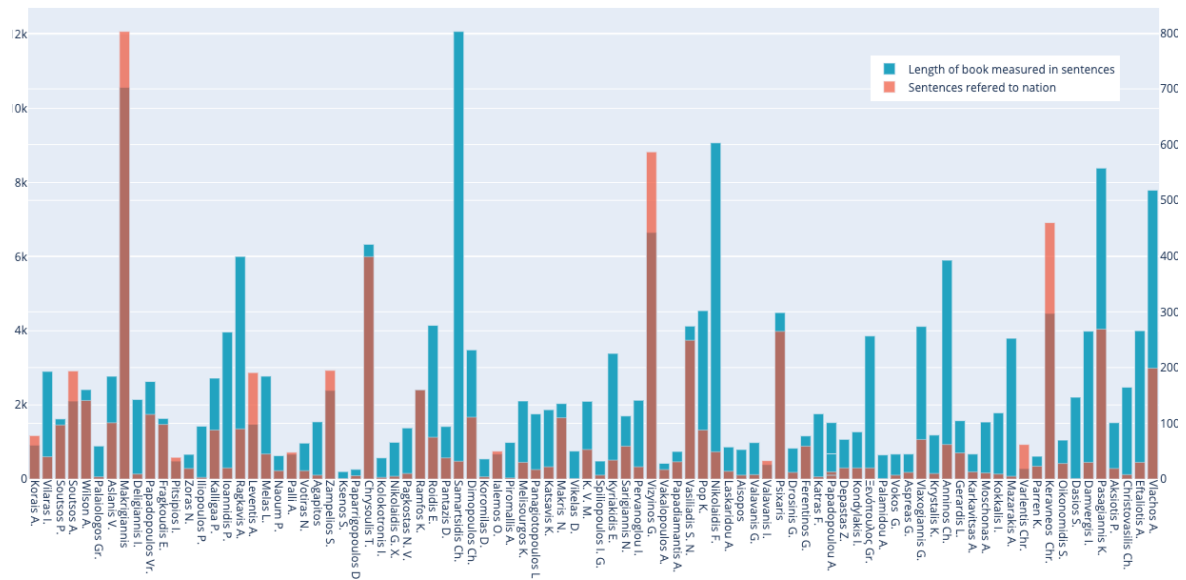
Figure 2: The sentences referred to nation and the total length of each book in sentences.

the reference rates per publication date. Figure 4 displays the discussion about the concept of nation in the corpus being more active and intense during the last two decades of the century. Figure 4, which takes into account the length of the sources, offers a more complete and objective look into the presence of the concept of the nation in the corpus we built.
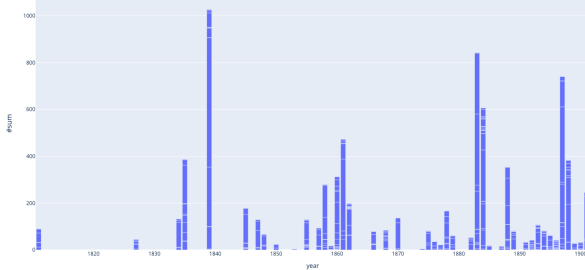

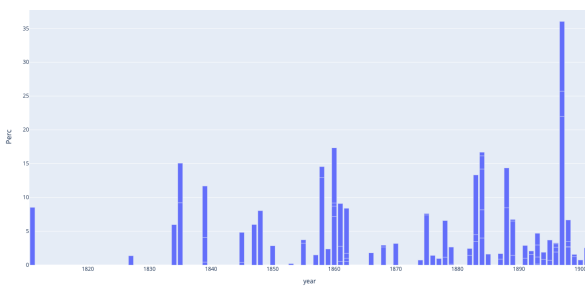Figure 3: The nation-related tokens in a timeline

Figure 4: The rate of nation-related sentences in a timeline

5.2 The presence of the keywords in the corpus

By elaborating further on the retrieved data, we could obtain information about the presence and frequency of each defined keyword. We decided to count the number of tokens retrieved by each keyword to get a better insight into the presence of the concept of nation in the body of 19th-century Greek fiction. This enabled us to observe the frequency of the defined keywords in the corpus across the 19th century. As we can see in the line chart of Figure 5, keywords such as "Hellas", "homeland" are by far most frequent, unlike keywords such as "tribe", "genus", "Romios"[6] and "Graecòs"/"Graecòs"[7] which are less frequently used in our corpus. It is very interesting that the national adjectives "Romios" and "Graecus" show very low numbers of occurrences and disappear at certain periods. For instance, the national adjective "Romios" shows the smallest frequency of occurrences among the keywords retrieved and it does not appear before 1839, as it is absent from the first seven documents of the corpus. More interestingly,

---

[6]"Romios" or "Romaios" was called the Orthodox Christian citizen under the Ottoman Empire. From the beginning of the 19th century, the term "Romios/Romaios" becomes equivalent to the term "Graecus"/"Graecòs" and declare the social and national identity of the Greek.

[7]The term "Graecus"/"Graecòs" which, from the Renaissance onwards retains the wider meaning of the speaker of the Greek language, during the 18th century declares the inhabitant of southeastern Europe who speaks Greek, and from the beginning of the 19th century is identified with the Greek ethnicity and means the Greek.

the national adjective "Graecus" does not appear in our corpus after the year 1888, as it is completely absent from 31 documents published between 1889 and 1901. The absence of the national adjective "Graecus" from 1888 onwards could be related to the general tendency observed in non-literary texts to avoid both "Graecus" and "Romios" and to use "Ellin" widely (Katsiardi-Hering, 2018).

5.3 The context of the concept of the nation

Based on the above-mentioned statistic analysis of the frequency and the dispersion of the concept of nation in our corpus, we continued by maintaining a top-down approach since we next focused on the context of the retrieved tokens. The main objective here was to obtain a more detailed picture of how the concept of nation was rendered and expressed in the 19th-century literary corpus we built. What was the vocabulary with which the concept of the nation is related and defined? What other concepts was it associated with?

Since there are no Part-of-Speech (POS) taggers for 19th-century Greek language, which would provide us with information about the syntactically related words to the tokens retrieved, we performed a collocation analysis. Collocation analysis revealed the context in which the concept of nation is shaped and defined in our corpus. The collected collocations were visualised with word clouds (also known as a tag cloud) which offered us a practical and compact insight into the context of each keyword individually and all the keywords as a whole. For instance, the 40 words that collocate most frequently with the keyword "nation" are shown in Figure 6.

As can be seen in Figure 6, of all the keywords we defined, only "Greece" (including the national adjective Greeks) and "homeland" appear to be associated with the keyword of the "nation". In fact, the token "Greece" is the most common collocation of the word "nation", and this indicates that the references to the term "nation" in our corpus refer primarily to the Greek nation. Except "Greece", the other geographical name that we can see on the cloud is the word "Europe" (upper right corner), which is an evidence of the strong political and ideological dimension Europe had for the construction of the modern state of Greece. However, some of the fundamental concepts of the Greek national integration seem to be condensed here in the high-frequency collocations of the keyword "nation", that include words such as "Greece", "today",

"justice", "fathers", "homeland", "freedom". Therefore, the issue of nation-building appears here as the urgent and topical demand for the restoration of justice and freedom. Furthermore, "fathers" that is a common Greek metaphor for the ancestors, is a central component of the national identity as it links together the past with the "today" and the "future" of the nation.

On the other hand, Figure 7 summarises the 40 words found in our corpus to collocate most frequently with all the seven keywords we defined ("nation", "Greece", "homeland", "genus", "tribe", "Graecus", "Romios"). The word cloud depicted in this Figure does not include the keyword-stem tokens. Even though it was usual for a keyword to be in a common context with other keywords (ie. in Figure 6 we saw "nation" to collocate with "Greece"), we use here a keyword-free cloud in order to observe their common context without any interference.

Interestingly enough, one of the high-frequency tokens of the word cloud of Figure 7 is the word "children", which is a metaphor for the citizens, since in phrases like "children of Greece" the to mean the people of Greece. In view of that, we can observe that this cloud reveals concepts that together emphasise the struggles for national independence and collectivisation, which is summarised in words like "today", "life", "blood", "land", "people", "freedom", "Union", "we" and "others". At the same time, the high-frequency tokens of the cloud are concepts like "today", populace ("children" and "people"), "blood" "life" and "name", which reflect the existential importance of the nation-building process. The notions of today and the populace could once again indicate how current and urgent the issue of nation-building was, while the blood, life, war, and guns could highlight the importance of the issue posed in between life and death. What is at stake in the struggle for national Independence is life itself. The token "name" is directly connected with the national designation, whether it refers to the collective name of the people or the personal name of a character.

Furthermore, both word clouds include concepts that are related to components of the national identity and which are dominant in shaping the national and collective consciousness. Concepts like these are freedom, land, language, ancestors (i.e. fathers), religion and God, and especially as regards Greece, honor, soul, glory and heart.
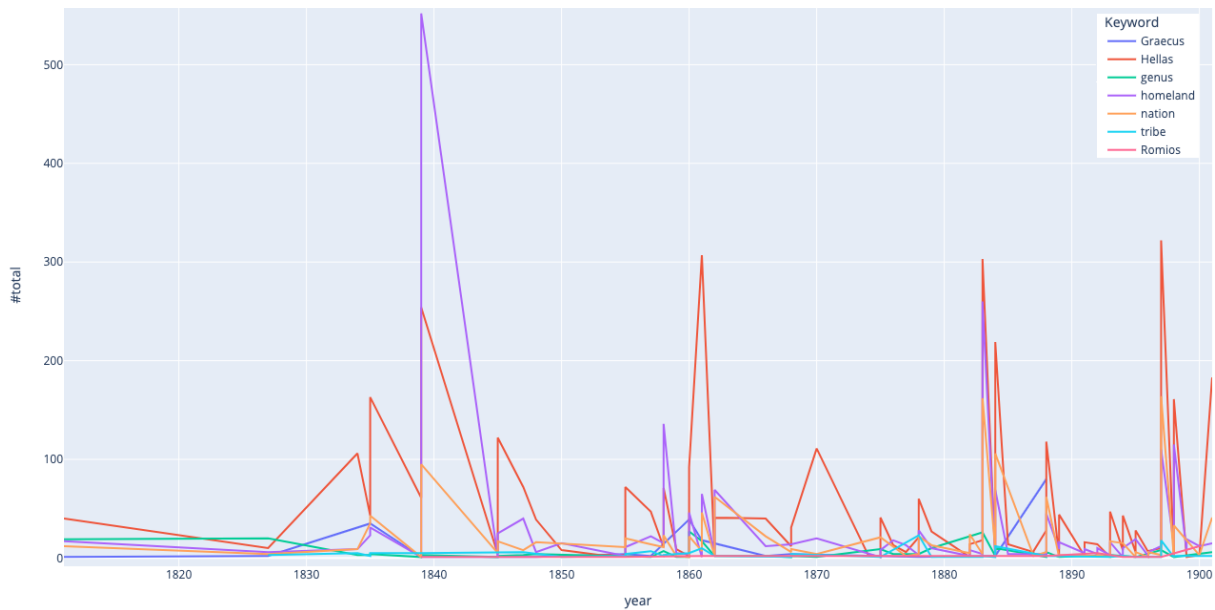
Figure 5: The keywords frequencies on a timeline.
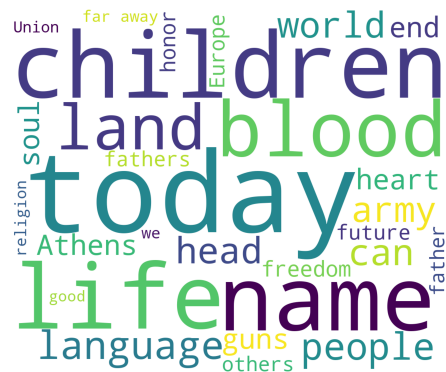

Figure 6: Collocation of "nation" keyword.

Figure 7: Collocation of all the seven keywords (nation, Greece, homeland, genus, tribe, Graecus, Romios).

5.4 The topic of nation-state

The unsupervised topic modeling analysis allowed us to automatically discover the thematic structures found within the entire corpus of the 19th-century Greek prose fiction. Even though this analysis was useful for the initial exploration of the corpus, we further exploited the topic modeling result by identifying the link between the nation theme and the extracted topic contents.

Out of the fourteen topics we extracted, we can see in Figure 8 that only the four most prevailing formed distinctive clusters. On the other hand, the ten least prevailing topics formed overlapping clusters, each of which was observed to contain heroes' names and was therefore associated with an individual novel. Likewise, the fourth topic, found in the right bottom corner of the figure, was observed to include vocabulary referring to the plot

and the characters of a particular novel, which has a unique theme, different from the rest of the corpus. However, the three most distinctive topics were associated with more general literary worlds, recurrent in our corpus. The first was comprised of terms related to the family relationships as well as to the religion (terms: god, head, heart, husband, daughter, blood, father, name, etc). The second topic included terms referring to the countryside, youth, and education (terms: home, child, village, language, brother, teacher, nature, etc). Finally, the third topic (in red circle) was observed to be highly associated with the concept of nation, since, as discussed in further detail below, it included terms referring to the State of Greece and its 19th-century history.
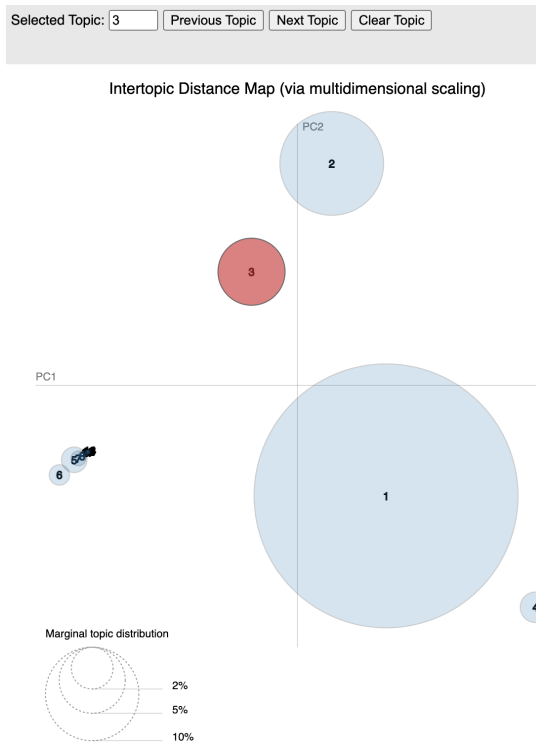
The third topic is considered to be the only one

Intertopic Distance Map (via multidimensional scaling)

Marginal topic distribution

- 2%
- 5%
- 10%

Figure 8: The topic which is referred to the nation-state of Greece.

Slide to adjust relevance metric:(2)

λ = 1   0.0   0.2   0.4   0.6   0.8   1.0

Top-30 Most Relevant Terms for Topic 3 (5.1% of tokens)

- Overall term frequency
- Estimated term frequency within the selected topic

1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

Figure 9: The topic which is referred to the nation-state of Greece.

that is directly and almost completely related to the concept of nation 9). In fact, the group of words that consists the cluster of the third topic seems to compose a condensed narration of 19th-century history of the newborn nation-state of Greece. For this topic, the most relevant term (given λ = 1) is the noun "homeland" which, at the same time, is the highest frequency keyword among those that we have defined as suggestive of the concept of the nation. Out of the thirteen nouns constituting the topic, the nine are terms already found in the collocation analysis of the nation keywords, such as "homeland", "man", "people", "others", "God", "Turk", "King", "land", and "war". The verbs that appear among the nouns constitute common enough verbs such as "do", "say", "come", "go", or "get" indicating the action of the nation-building progress.

By reducing the relevance value below 0.5, we observe the formation of word groups that refer to either the government establishment or the liberal-national revolution. For example, by setting the value of λ at 0.4 we get a word group such as "caste", "government", "people", "governor", and "homeland that seems to reflect the issue of governance in the newborn state of Greece. While, at 0.25 (and below that) the words "ammu-

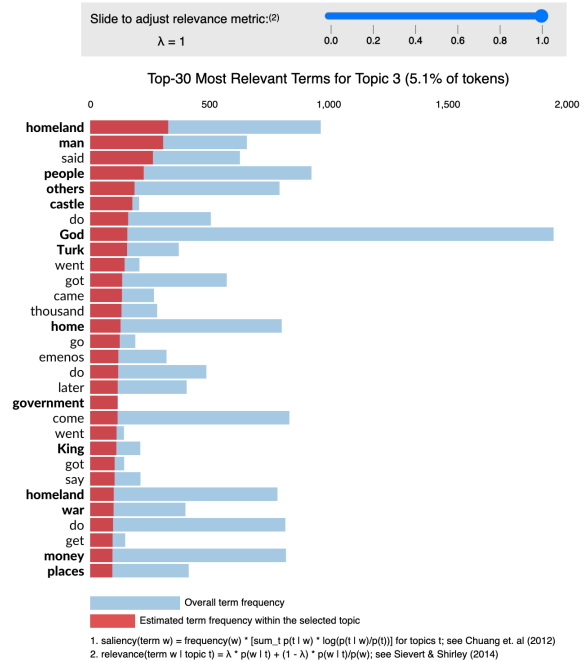nition", "killed", "people", "Kolokotronis", "warriors", "fought", "patriots", "drachma" (coin currency), and "elections" are added at the bottom. Both word groups together may form a direct reference to the issue of the warriors' demand to participate in the governance of the free state of Greece. Considering the different topic contents produced by the multiple relevance adjustments, we may say that the third topic is directly connected with the discourse about the establishment of the first Hellenic Republic and the formation of the modern state of Greece. Even though the topic contains traces of the nation theme, it is clear that it also extends to statehood issues like governance or economy.

6 Conclusions

Considering the challenge of the 19th-century Greek language, computational analysis of prose fiction offered us valuable insights into the special issue of the construction of the national identity. Using the keyword-based retrieval technique we managed to track down all the direct references to the concept of the nation. The statistical analysis of these findings allowed us to observe the size and the frequency of the nation-relate discussion across our corpus and its timeline. In order to produce accurate and realistic measurements, we used metadata about the length of each document that

enabled us to estimate the rates of nation-related references across the corpus. As an overall observation on the presence of the concept of nation in our corpus, it was noticed an increasing and escalating tendency during the last two decades of the century.

Focusing on the phraseology of the nation-related references, we also attempted to identify the vocabulary through which the concept of the nation is expressed and shaped. By counting the retrieved keywords, we confirmed the Greek peoples' preference -witnessed by non-literary sources-to self-determine themselves as *Hellines* (singular *Hellin*) rather than *Romios* or *Graecus*. Moreover, the keywords we defined as direct references to the nation, were further analyzed with a focus on their collocations. The collocations extracted reflect aspects of the Greek national identity and milestones of the Greek history of the 19th century related either to the War (of Independence) or the challenges of the nation-building process.

However, the collocations' word clouds may provide a slightly different picture of that we get by analysing the topic modeling results. In fact, the results of the collocation analysis were found to be associated with one of the thematic patterns that have the highest probability in the corpus. This topic, which is the third words' cluster in the topic modeling analysis, brings together terms that are linked to the nation and refer, as mentioned above, either to the War of Independence or to the institutionalisation and the governance. Considering the several versions that the topic may form by the adjustments of the relevance value, we notice that the concept of the nation is present in a topic that, at the same time, reflects the pragmatic challenges of the state-building process (i.e. government, governor, king, money, elections). At the same time, the collocations bring together terms that make up the picture of romantic nationalism, since they combine concepts like justice, idea, freedom, glory, honor, soul, heart, life, hope, belief, Union.

The methodological approach described above enabled us to observe the concept of the nation in the corpus of 19th-century Greek prose fiction we built. However, the result, as well as the method, could be further enriched and developed both by extending the content of the corpus and by developing a 19th-century specific language toolkit, which would provide us with more detailed and comprehensive insights into the 19th century Greek Literature.

Acknowledgments

References

Roderick Beaton. 2021. The significance of the 1821 revolution for greece and the world. *ekathimerini-com*.

Roderick. Beaton and David. Ricks. 2009. *The Making of Modern Greece : Nationalism, Romanticism, and the Uses of the Past (1797-1896)*. Taylor Francis Ltd.

David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022.

Earl L. Bradsher. 1921. Nationalism in our literature. *The North American Review*, 213(782):109–118.

Matt Erlin, Andrew Piper, Douglas Knox, Stephen Pentecost, Michaela Drouillard, Brian Powell, and Cienna Townson. 2021. Cultural capitals: Modeling 'minor' european literature. *Journal of Cultural Analytics*, 1(2):21182.

John H. Fisher. 1980. Literary culture: Nationalism and the study of literature. *The American Scholar*, 49(1):105–110.

Mary E. Giffin. 1945. Nationalism and english literature. *College English*, 6(6):310–313.

Olga Katsiardi-Hering. 2018. Ellin, graikos, romios: from multinational to national. In *Ellin, Romios, Graikos: Collective Identitifications and Identities*, volume 7, chapter 1, pages 19–35. Evrasia, Athens.

Alexis Politis. 2017. *Romantic Literature in the National State*. Philologia. Crete University Press, Herakleion.

Efstathios Pouliasis. 2021. *The perception of the Greek Revolution in the Public Sphere of the GreekState (1832-1920). Dissertation*. Ionian University: Faculty of History Translation, Corfu.

Radim Rehurek and Petr Sojka. 2011. Gensim–python framework for vector space modelling. *NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic*, 3(2).

Anna Tampaki and Ourania Polikandrioti. 2016. *Greekness and Diversity; Cultural Mediation and National Character in Nineteenth Century*, volume A B. Athens: National and Kapodestrian University of Athens National Research Institute, Athens.

Dimitris Tziovas. 2017. *The Cultural Poetics of Greek Fiction: From Interpretation to Ethics*. Philologia. Crete University Press, Herakleion.

Maartje Weenink. 2018. *Who's Afraid of the Big Bad?: The Representation of Nationality in "British" Gothic Fiction 1750-1840. A Computational Approach to Topics in Fiction. Thesis*. Radboud University: Department of Historical, Literary and Cultural Studies, Nijmegen, Netherlands.