

Deniz T. Kılınçoğlu and Jörg Wettlaufer

The Reshuffling of Middle Eastern Identities in the Age of Nationalism: Insights from Nineteenth-Century Travelogues

Travelogues about the nineteenth-century Middle East are a rich source of information about the material culture, economic life, and social practices and customs of various religious and ethnic communities of the region. In addition, many European travelers provide us with colorful observations about inter-communal relations and sentiments. These accounts reflect the changing ideas and emotions about the “self” and the “other(s),” at a time when newly emerging national identities gradually overshadowed previous (mostly religious and local) dominant forms of self-identification. In this respect, travelogues contain first-hand observations about the sociopolitical dynamics of the rise of nationalisms and the consequent identity (re-)formation in the nineteenth-century Middle East.

However, many such accounts suffer from some well-known problems of this genre. We observe, for example, that the authors’ own (mostly white, male, Christian, European) self-identification shapes their perceptions of the identity dynamics in the region. Most of these works reflect the Orientalist *Zeitgeist* of the age, openly favoring Christian, rather than Muslim, perspectives on social, political, and cultural issues. Moreover, we occasionally find that some of these accounts are (at least partly) copied from earlier sources, thereby not reflecting first-hand observations and experiences. An even bigger problem is that some of these texts are purely fictional narratives produced merely as sensational products for the market. These potential (and occasionally serious) pitfalls in using this genre as source have caused historians to approach this large body of historical documents with understandable caution. Nevertheless, in the present chapter we argue that travelogues are invaluable sources for historical research,¹ especially with the help of the recently developed analytical tools of digital humanities.

¹ For earlier debates on travelogues as potential historical sources, despite all the problems mentioned here, see Barbara Korte, *English Travel Writing: From Pilgrimages to Postcolonial Explorations* (Basingstoke 2000), pp. 5–16; Daniel Carey, *Truth, Lies and Travel Writing*, in Carl Thompson (ed.), *The Routledge Companion to Travel Writing* (London 2015), pp. 3–14.

This study reflects some preliminary observations resulting from our digital history project that aims at building and digitally analyzing an extensive corpus of English-, French-, and German-language travelogues about the nineteenth-century Ottoman Empire.² It discusses our initial findings in travelogues regarding the sociopolitical dynamics of the rise of Middle Eastern nationalisms and our methodological suggestions about the opportunities as well as challenges of employing digital tools to use travelogues for historical analysis. The chapter starts with a discussion on nations and nationalisms in the nineteenth-century Middle East, particularly in the Ottoman Empire. Then it deals with the question of using travelogues as historical sources, focusing on the history of nationalisms, and suggests that digital history tools are essential for increasing the reliability of this body of sources. Finally, we present our findings and suggestions on how and what kind of digital history tools can be used to open such corpora to historical research.

1 Nations and nationalisms in the Ottoman Empire

The nineteenth century was the age of great transformation for the Middle East as well as for the Western world. The pace of intellectual, cultural, economic, political, and social change was dazzling compared to earlier centuries. As the region opened itself to new ideas, concepts, and institutions, especially from Europe, these novelties inspired endeavors and plans for social, economic, and political change to adapt local institutional frameworks to the new market-oriented, industrial, and science-and-technology-led global world-system.

One of the novelties of the modern world-system was a new doctrine that had been reshaping the political and diplomatic institutional framework and common sense especially since the beginning of the century. This doctrine of European origin was called “nationalism.” It held simply “that humanity is naturally divided into nations, that nations are known by certain characteristics which can be ascertained, and that the only legitimate type of government is national self-government.”³ Adapting to this new doctrine and its ‘international’ institutional framework was a great challenge for the empires that rose on multi-ethnic sociopolitical institutions. Eventually, the long nineteenth century ended with

² For more information about the project see *Travels in the 19th-Century Ottoman Empire*. A Digital History Research Project, accessed 4 April 2022, <<http://middle-east-travelers.de/>>.

³ Elie Kedourie, *Nationalism*, 2nd and revised ed. (London 1961), p. 9.

several empires collapsing and dissolving into nation-states or transforming themselves into nation-state-like formations. In any case, the following century would be the age of nations and nation-states. The Ottoman Empire was one of the collapsing empires. Its dissolution gave way to many nations and nation-states in the Balkans and the Middle East, leaving a lingering legacy in the economic, political, and cultural lives of its (mostly reluctant) inheritors in the region.

Starting from the early nineteenth century, the elites of the region, both of the imperial center and of various ethnoreligious provincial communities, tried to adapt to the nationalism-dominated new world order. In the process, they had to reinvent themselves and the sociopolitical institutional frameworks that they led and represented. This paved the way for various centrifugal and centripetal forces, and naturally, occasional clashes between these forces. The main centrifugal force in this sense was the rise of nationalist ideas and movements starting in the Balkans in the early nineteenth century. The central imperial elites tried to counter these endeavors by various potentially centripetal measures, from political reforms that granted equal citizenship rights to all ethnoreligious communities to violent military measures against nationalist uprisings. Nevertheless, they could not stop nationalist ideas and movements gaining more ground in the Empire, which culminated in its dissolution into nation-states at the end of the First World War.⁴

There is a now relatively rich literature on the rise of nations and nationalisms in the Middle East and the Balkans in the long nineteenth century. Many scholars have studied this question focusing on individual cases, such as Greek, Armenian, Turkish, and Arab nationalisms.⁵ Following the dominant trends in nationalism studies in general, most of these works focus on the intellectual, political, and diplomatic aspects of the question.⁶ Thanks to these studies, we have a relatively strong knowledge base about nationalist thinkers and

4 For a concise analytical account of the last Ottoman century, revolving around these issues, see M. Şükrü Hanioglu, *A Brief History of the Late Ottoman Empire* (Princeton 2008).

5 Just to name a few examples: Hasan Kayalı, *Arabs and Young Turks. Ottomanism, Arabism and Islamism in the Ottoman Empire, 1908–1918* (Berkeley et al. 1997); Rashid Khalidi et al. (eds.), *The Origins of Arab Nationalism* (New York 1991); Michael Provence, *The Great Syrian Revolt and the Rise of Arab Nationalism* (Austin 2005); Keith David Watenpugh, *Being Modern in the Middle East. Revolution, Nationalism, Colonialism, and the Arab Middle Class* (Princeton, Oxford 2006).

6 For two reviews of this literature see Miroslav Hroch, *National Movements in the Habsburg and Ottoman Empires*, in John Breuilly (ed.), *The Oxford Handbook of the History of Nationalism* (Oxford 2013), pp. 175–198; Aviel Roshwald, *Nationalism in the Middle East, 1876–1945*, in Breuilly, *Oxford Handbook of the History of Nationalism*, pp. 220–241.

their ideas, and about the role of politics and diplomacy on the realization of these ideas into post-Ottoman nations and nation-states. In other words, the existing literature explains how the elites (e.g., intellectuals and the bourgeoisie) of these different Ottoman communities and international diplomacy played significant roles in reshaping the region in the nineteenth and early twentieth centuries.

Despite some introductory efforts,⁷ however, we are still far from having a satisfactory understanding of the social history of the rise of nationalisms in the Ottoman Empire. Hence, we still need analytical and comparative studies on the question of how inter-communal relations in the Ottoman Empire paved the way for and shaped post-Ottoman nations and nationalisms in the region. As some researchers have mentioned,⁸ mutual emotions between the Ottoman communities influenced the early formation of nationalisms in the empire, which in turn shaped the post-Ottoman formulations of nations and nation-states in the region.

As emphasized by many scholars of nationalism, certain shared narratives in the forms of “myths, symbols, and memories” constitute vital elements of nationalism and of the feeling of nationhood.⁹ Many of these elements are inherited from pre-nationalist cultural arsenal of ethnoreligious communities that dominate the cultural formation of modern nations.¹⁰ In this regard, it is crucial to examine pre-national ethnoreligious sentiments of belonging and the intercommunal tensions and relations between these groups, which are more often than not translated into the realities of inter-ethnic and international relations in the age of nationalism. Ego documents from the age of the rise of nations and nationalisms, in this respect, provide us with unexplored areas of first-hand knowledge about these issues. Memoirs, autobiographies, and other forms of eyewitness accounts from the nineteenth-century Middle East stand out naturally for the research on the experiential history of nationalism in the region. However, eyewitness travel accounts about the nineteenth-century Middle East that are written by Europeans should also be considered in this respect as they enrich

7 See, for example, Fatma Müge Göçek (ed.), *Social Constructions of Nationalism in the Middle East* (Albany 2002).

8 M. Şükrü Hanioglu, *Turkish Nationalism and the Young Turks, 1889–1908*, in Göçek, *Social Constructions*, pp. 85–97; Fatma Müge Göçek, *Decline of the Ottoman Empire and the Emergence of Greek, Armenian, Turkish, and Arab Nationalisms*, in Göçek, *Social Constructions*, pp. 15–83.

9 For a detailed treatment of the subject see Anthony D. Smith, *Myths and Memories of the Nation* (Oxford 2000).

10 Anthony D. Smith, *The Ethnic Origins of Nations* (Malden 1986).

our understanding of the ethnoreligious formations and inter-communal tensions in the region with their outsiders' perspectives.

2 Travelogues as a source for the history of nationalisms in the nineteenth-century Middle East

Travelogues about the nineteenth-century Middle East are a rich and still largely untapped source for historical research. They provide researchers with abundant information about the economic, cultural, social, political, and even intellectual life in the region in this tumultuous age. However, some well-known problems of the genre led historians to be highly skeptical about these sources. Moreover, the scattered and unstructured form of this important literature hinders systematic and comprehensive use of the genre. We believe, however, that the historians' methods, rather than the nature of the documents, will be determinant in turning this rich and interesting body of literature into more reliable sources for historical research.

The main reason for the historians' relative reluctance to use this body of literature as a historical source, and therefore the relative poverty of systematic research on this genre, is the "subjectivity" of these documents. Travel literature is indeed a "hybrid genre containing autobiographical, novelistic, essayistic, and journalistic elements"¹¹. We observe, in most of these texts, that political, cultural, emotional, and stylistic concerns are interwoven with the observed "facts." This has led many historians to regard these texts as unreliable sources for research. Historical research on travelogues, as a result, has mostly focused on the travelers themselves, especially their perceptions of other cultures and political assumptions, rather than the material information that their travel writings contained.¹² Following this format, a currently running digital history project on

11 Elisabeth Oxfeldt, *Journeys from Scandinavia: Travelogues of Africa, Asia, and South America, 1840–2000* (Minneapolis, London 2010), p. xxi.

12 Andrew Hammond, *The Uses of Balkanism: Representation and Power in British Travel Writing, 1850–1914*, in *The Slavonic and East European Review* 82/3 (2004), pp. 601–624; Iain Macleod Higgins, *Writing East: The "Travels" of Sir John Mandeville* (Philadelphia 1997); Geoffrey P. Nash, *From Empire to Orient: Travellers to the Middle East, 1830–1926* (London 2005); Paul Manning, *Strangers in a Strange Land: Occidentalists Publics and Orientalist Geographies in Nineteenth-Century Georgian Imaginaries* (Boston 2012), pp. 28–58; Barbara Spackman, *Accidental Orientalists: Modern Italian Travelers in Ottoman Lands* (Liverpool 2017).

German-language travelogues focuses particularly on self-representation and otherness in these texts.¹³

We claim that despite their obvious problems, the travelogues are invaluable sources for historical research. It is imperative to remember at this point, regarding especially the issue of “subjectivity” in travel writing, that almost all primary sources that historians use in their research are subjective accounts. In other words, dismissing eyewitness travel accounts as subjective, and therefore unreliable, reflects an absurd scholarly assumption about the putative existence of “objective” historical sources that show us “how it actually was” (“*wie es eigentlich gewesen ist*”) as Leopold von Ranke (1795–1886) once optimistically assumed. What makes this assumption rather dangerous is to attribute such a nature to more conventional sources, such as official documents or court registers, which obviously were shaped by certain political, ideological, and even economic assumptions and concerns. In any case, it is the historians’ methods, more than the nature of these documents, that will reveal the useful content in these rich sources. As historians adopt a critical, analytical, and comparative approach to these narratives (such as by asking what these narratives tell, who composed them and why, and by comparing the contents with other similar documents), they will develop powerful methods to integrate these sources into their research. In this regard, recently developed tools and methods in the digital humanities will provide historians with previously unavailable scholarly command of these and other similar sources. These tools will especially be useful in two major issues of the literature, which render travelogues slightly more unreliable compared to other sources in the eyes of the historians: first, in separating fact and fiction, and second, in addressing copying within the genre. As discussed below, digital text analysis provides us with powerful tools to detect these problems and separate useful historical content from copied texts and authorial fantasies.

Travelogues about the nineteenth-century Middle East contain abundant information on material culture, economic life, social practices and customs, and biographical information about many urban and rural notables of the era. This is the reason why many historians have used these sources for their research.¹⁴ Yet,

13 Travelogues: Perceptions of the Other 1500–1876 – A Computerized Analysis, accessed 18 May 2022, <<https://travelogues.github.io/>>.

14 Just to name a few examples among many: Benjamin Braude and Bernard Lewis (eds.), *Christians and Jews in the Ottoman Empire: The Functioning of a Plural Society* (2 vols., New York 1982); Christopher G. A. Clay, *Gold for the Sultan: Western Bankers and Ottoman Finance 1856–1881* (London 2000); Suraiya Faruqi and Christoph K. Neumann (eds.), *Ottoman Costumes: From Textile to Identity* (Istanbul 2004); Deniz T. Kılınçoğlu, *Economics and Capitalism*

other difficulties in working with travelogues arise at this point. First, there is no systematic corpus or even a complete detailed bibliography of this genre.¹⁵ Second, many of these sources are either not digitalized or the digital copies are scattered across various digital vaults. Therefore, historians must work manually to find relevant information for their research. This mostly leads them not only to spending much time and energy on finding useful information, but also to overlooking many relevant sources and their valuable content. These problems can be solved only by building a comprehensive and systematic corpus of travelogues and a powerful online portal that will provide the historians with various tools to extract information from these rich sources in organized and visualized ways.

Coming back to the question of using eyewitness travel accounts for understanding the rise of nationalisms in the region, many authors of this genre provide us with their observations about the inter-communal relations among various ethnic and religious communities of the region. Moreover, these accounts include detailed observations about the everyday life of various ethnoreligious communities, relationships between them, and some remarks on rising tensions resulting from newly emerging national(ist) sentiments. These observations about everyday life are enriched, in some travelogues, by visual representations of various cultural aspects of the ethnoreligious groups as well as depictions of scenes from everyday life.¹⁶ Another element in some of these accounts that make them attractive sources for historians is their inclusion of conversations with local (naturally, mostly elite) people. An important example in this context

in the Ottoman Empire (London 2015); Baki Tezcan, *The Second Ottoman Empire: Political and Social Transformation in the Early Modern World* (Cambridge 2010); Ehud R. Toledano, *Slavery and Abolition in the Ottoman Middle East* (Seattle 2012).

¹⁵ There have been earlier attempts to produce such bibliographies. However, they mostly focus on pre-nineteenth-century works. See, for example, Elisabetta Borromeo, *Voyageurs occidentaux dans l'Empire ottoman (1600 – 1644): Inventaire des récits et études sur les itinéraires, les monuments remarquables et les populations rencontrées (Roumélie, Cyclades, Crimée)* (Paris 2007); Julia Chatzipanagioti-Sangmeister, *Griechenland, Zypern, Balkan und Levante: Eine kommentierte Bibliographie der Reiseliteratur des 18. Jahrhunderts* (2 vols., Eutin 2006); Stéphane Yerasimos, *Les voyageurs dans l'Empire Ottoman (XIV^e–XVI^e siècles): Bibliographie, itinéraires et inventaire des lieux habités* (Ankara 1991).

¹⁶ For a few interesting examples see Peter Edmund Laurent, *Recollections of a Classical Tour through Various Parts of Greece, Turkey, and Italy. Made in the Years 1818 & 1819*, 2nd ed. (2 vols., London 1822); Luigi Mayer, *Views in the Ottoman Dominions, in Europe, in Asia, and Some of the Mediterranean Islands* (London 1810); G. A. Olivier, *Travels in the Ottoman Empire, Egypt, and Persia, Undertaken by Order of the Government of France, During the First Six Years of the Republic* (2 vols, London 1801); [Julia Sophia H.] Pardoe, *The Beauties of the Bosphorus [...]* (London 1838); [Jerome] V[an] C[rowninshield] Smith, *Turkey and the Turks* (Boston 1854).

is the travel account of the well-known British economist and the first professor of political economy at Oxford (1825–1830), Nassau William Senior (1790–1864). Senior, in his *A Journal Kept in Turkey and Greece in the Autumn of 1857, and the Beginning of 1858*, recounts many conversations with Ottoman elites as well as European observers who resided in the Ottoman Empire at the time about the social, economic, and political conditions in the empire.¹⁷ These conversations reveal not only first-hand observations about various issues in the Ottoman realm but also ideas, disputes, and emotions of the contemporary observers regarding these issues. In this sense, these accounts constitute an invaluable source for historians working in a wide range of fields from economic and diplomatic, to intellectual and cultural, history.¹⁸

European authors' observations reflect obviously subjective evaluations (mostly, for example, in favor of the Christians against Muslims) in most of these accounts. Nevertheless, they still provide us with significant insights into the social, political, and cultural settings in which Middle Eastern nationalisms emerged in relation to one another. In this regard, we also find occasional references to mutual emotions and other manifestations of rising (proto)nationalist tensions among these communities. It will suffice to give a few examples at this point, leaving a more comprehensive analysis to prospective digital research on the entire corpus.

The English traveler Charles Boileau Elliott (1803–1875) observes the following in the 1830s about Kastellorizo (Castellorizo, Castle Rosso, or Meyisti), which is now a Greek island that lies two kilometers off the south coast of Turkey:

The defences of Castel Rosso are two forts, both in a state of ruin, one of which is tenanted by a Turk and his wife, the latter of whom greeted us unveiled [...]. The population consists of about three thousand Greeks; to whom a hundred Turks from the continent have hitherto been added to represent the Ottoman interests; but as their number is so small, they assume none of the airs of superiority peculiar to their nation, and even adopt Greek customs, as was strikingly evinced in the case of the unveiled female. A day or two before our arrival an order was received from Constantinople that these few Turks should vacate the island, which is henceforth to be consigned exclusively to the Greeks on payment of a fixed tribute. The rejoicing on this occasion was unbounded. Young and old seemed to partake the senti-

¹⁷ Nassau W[illiam] Senior, *A Journal Kept in Turkey and Greece in the Autumn of 1857 and the Beginning of 1858* (London 1859).

¹⁸ For a few examples of such conversations between Senior and his Ottoman and European interlocutors about the economic conditions and policies in the Ottoman Empire see Kılınçoğlu, *Economics and Capitalism*, pp. 19, 25–26, 130.

ment; even our grey-headed captain, a native of the rock, forgetful of his age and his cares, sang and danced with joy, evincing all the gaiety and light-heartedness of youth.¹⁹

Continuing on the theme of nationalist sentiments and mutual sentiments among ethnoreligious communities, the following words from the American statesman Samuel G. W. Benjamin (1837–1914) are particularly striking:

Another feature noticeable in the character of Constantinople society is the almost total absence of public spirit. This arises from the circumstance that, figuratively speaking, every man's hand is against the neighbor. In the capital, as throughout the empire, we see the people divided into conquerors and conquered, oppressors and oppressed, and a war of races is the natural result. [...] [T]he many nations of Turkey have guarded their individuality through long ages of misrule. As a consequence the Armenian regards the Greek with jealousy, and the Greek in turn regards the Armenian with distrust, each watching the lest the other acquire undue influence at the Porte, and both unite in cursing the Turk; while he hates them as Christian dogs who may eventually oust him from his dominion over them.²⁰

As shown in these examples, travelogues contain many first-hand (and often very striking) observations about the sociopolitical dynamics of the rise of nationalisms and the consequent identity (re-)formation in the nineteenth-century Middle East. The job of the historian is not to take these observations at face value, but to analyze them critically and comparatively, using both other primary sources and relevant scholarly analyses. In this regard, the tools offered by the digital humanities provide unrivaled opportunities to sift through these texts to reveal relevant information and to employ intertextual analysis (with other travelogues as well as alternative historical sources) to analyze the information in historical context.

The first step toward turning travelogues into useful historical sources, we believe, is to build large-scale corpora that are optimized for global search options and large-scale intertextual analyses. Especially the latter will help us see intertextual borrowings that reveal both influences and copying. Detecting influences is important for intellectual historians wishing to observe continuities and change in perceptions about the region. However, picking out copying is especially critical in separating genuine observations from fictional and derivative ones. This is indispensable for building a reliable corpus that historians can use

19 C[harles] B[oileau] Elliott, *Travels in the Three Great Empires of Austria, Russia and Turkey* (2 vols., London 1838), vol. 2, pp. 199–201.

20 S. G. W. Benjamin, *The Turk and the Greek; or, Creeds, Races, Society, and Scenery in Turkey, Greece, and the Isles of Greece* (New York 1867), pp. 30–31.

for their research on the region. Since “[t]wo particular modes of writing – forgery and its respectable cousin, parody – have specially close, even parasitic, relationships with travel writing,”²¹ scholars of the genre have always suffered from problems of authenticity in travel writing.²² We believe that digital methods and tools for textual analysis will finally allow historians to minimize such problems, enabling them to explore this rich body of historical sources with more confidence.

3 Digital research tools and resources for travelogues

There are only few digital resources and online portals that can enable researchers to work with nineteenth-century travelogues about the Ottoman Empire. Digital libraries such as Google Books²³ or HathiTrust²⁴ do not offer the possibility of limiting text mining to predefined collections or to carry out more complex analyses such as topic modeling or navigation through the text with the help of named entities. Although some national libraries, such as the Bibliothèque Nationale de France, allow narrowing down the corpus to relevant collections within their extensive digitized resources, there is still no easy way to separate travels to the Ottoman Empire from the others.²⁵

21 Peter Hulme and Tim Youngs, Introduction, in Peter Hulme and Tim Youngs (eds.), *The Cambridge Companion to Travel Writing* (Cambridge 2002), pp. 5–6. For an analytical treatment of the interplay between “fictional and documentary texts on travel,” focusing especially on the literature about the Arctic, see Janicke Stensvaag Kaasa, *Travel and Fiction*, in Nandini Das and Tim Youngs (eds.), *The Cambridge History of Travel Writing*, (Cambridge 2019), pp. 474–487.

22 For a brief historical analysis of the “long tradition of suspicion about the veracity of travel writing,” see Daniel Carey, *Truth, Lies and Travel Writing*, in Carl Thompson (ed.), *The Routledge Companion to Travel Writing* (London 2015), pp. 3–14.

23 Google Books, accessed 18 May 2022, <<https://books.google.de/>>.

24 HathiTrust. Digital Library, accessed 18 Mai 2022, <<https://www.hathitrust.org/>>.

25 As an example, see Édouard Charton et al. (eds.), *Le Tour du monde, nouveau journal des voyages* (Paris 1860–1895, new series 1895–1914), accessed 18 May 2022, <<http://catalogue.bnf.fr/ark:/12148/cb32878283g>>. *Le Tour du monde* was a weekly journal that published travels accounts from all around the world. One way for narrowing down the corpus to travels in the Ottoman Empire is keyword search within the collection. Another has been described recently by Jan Rörden et al., *Identifying Historical Travelogues in Large Text Corpora Using Machine Learning*, in Anneli Sundqvist et al. (eds.), *Sustainable Digital Communities. 15th International Conference, iConference 2020. Borås, Sweden, March 23–26, 2020. Proceedings* (Cham 2020), pp. 801–815.

In the early 2000s, Ralph C. Müller prepared an exhaustive database on travelers and migrants in the Ottoman Empire between 1396 and 1611. Müller eventually published the database in ten hard-copy volumes and on CD-ROM.²⁶ In this regard, it represents a rather typical approach in the transitional period between non-digital and digital research tools. More recently, digital resources have become available online usually through thematic portals. For example, Jörg Wettlaufer started the portal *www.digiberichte.de* in 2006/2007 as part of a project supported by the Göttingen Academy of Sciences and Humanities.²⁷ It offers prosopographical and bibliographical information as well as access to full texts of European travelers of the late Middle Ages (1250 – 1530), primarily on pilgrimages to the Holy Land. However, the functionality in this portal is rather limited compared to contemporary standards. Only simple keyword in context (KWIC) queries are possible, and the plain texts are stored and queried in a relational database, which is no longer the state of the art. The portal has enjoyed great popularity in the scholarly community since its public release and is often used by researchers from Southeast Europe.

The travelogues project of the Greek Aikaterini Laskaridis Foundation focuses on the visual materials in travel reports about the Middle East and the Mediterranean from the fifteenth to the twentieth centuries.²⁸ The project website exhibits many examples of such graphic materials and provides access to valuable additional sources for this period. Another recent research project on travelogues using methods of the digital humanities and natural language processing is the

26 Ralph C. Müller, *Prosopographie der Reisenden und Migranten ins Osmanische Reich (1396 – 1611)* (Berichterstatter aus dem Heiligen Römischen Reich, außer burgundische Gebiete und Reichsromania) (10 vols., Leipzig 2006). Another source for this period is Hélène Pignot's *La Turquie chrétienne: Récits des voyageurs français et anglais dans l'Empire ottoman au XVII^e siècle* (Vevey 2007). For similar bibliographies see note 15 above.

27 The website builds on the following series of three volumes edited by Werner Paravicini between 1994 and 2001: *Europäische Reiseberichte des späten Mittelalters, Eine analytische Bibliographie*, vol. 1: *Deutsche Reiseberichte*, bearbeitet von Christian Halm (Frankfurt am Main 1994); vol. 2: *Französische Reiseberichte*, bearbeitet von Jörg Wettlaufer in Zusammenarbeit mit Jacques Paviot (Frankfurt am Main 1999); vol. 3: *Niederländische Reiseberichte*, bearbeitet von Jan Hirschbiegel (Frankfurt am Main 2000). For a detailed description of the project, see Jörg Wettlaufer, *Europäische Reiseberichte des Späten Mittelalters: Das Projekt einer Digitalisierung der Editionen und eines Themenportals im Internet*, in Bernard Guenée and Jean-Marie Moeglin (eds.), *Relations, échanges et transferts en Europe dans les derniers siècles du Moyen Age: Hommage à Werner Paravicini* (Paris 2010), pp. 539 – 555.

28 TRAVELOGUES – TRAVELERS' VIEWS. Places – Monuments – People. Southeastern Europe – Eastern Mediterranean – Greece – Asia Minor – Southern Italy. 15th–20th Century, accessed 18 May 2022, <<http://eng.travelogues.gr/>>.

“Travelogues” project located in Vienna and Hannover.²⁹ The emphasis here is on the semi-automatized collection and analysis of German language travelogues in the holdings of the Austrian National Library printed between 1500 and 1876. The thematic focus of the project is on the identification of intertextual relations among and the analysis of perceptions of “otherness” and “Orient” in the travelogues based on the (imperfect) full texts created by Google Books with the help of Optical Character Recognition (OCR) engines.³⁰

4 An online portal for research on travelogues to the Ottoman Empire³¹

As described earlier, we believe that effective research on large corpora of travelogues needs support from tools and methods developed within the digital humanities. Therefore, our interdisciplinary project is situated at the intersection of digital research and history. It proposes methods such as named entity recognition and aims to apply them to multilingual and historical corpora with the help of machine learning (ML) and neural networks in particular. Further project goals are the provision of a text mining environment and the establishment of linked open data for the identified entities and metadata within the travelogues.³²

29 TRAVELOGUES: Perceptions of the Other and the Orient in Modern Times: A Mixed-method Approach for the Analysis of Large-scale Travel Account Series, accessed 18 May 2022, <<https://travelogues.github.io/>>. On this project see Doris Gruber, Elisabeth Lobenwein, and Arno Strohmeier, *Travelogues – Perceptions of the Other 1500–1876: A Computerized Analysis*, in Thomas Wallnig, Marion Romberg, and Joëlle Weis (eds.), *Achtzehntes Jahrhundert digital: zentraleuropäische Perspektiven* (Vienna et al. 2019) (= *Das Achtzehnte Jahrhundert und Österreich. Jahrbuch der Österreichischen Gesellschaft zur Erforschung des Achtzehnten Jahrhunderts* 34), pp. 129–131; Doris Gruber et al., *Travelogues: Fremdwahrnehmungen in Reiseberichten 1500–1876*, in Marlene Ernst et al. (eds.), *digital humanities austria 2018: empowering researchers* (Vienna 2020), pp. 62–66, accessed 18 May 2022, <<https://epub.oeaw.ac.at/?arp=0x003b3d14>>. See also Rörden et. al. cited in note 25.

30 The project is using the platform Recogito, developed by the Pelagios consortium (accessed 18 May 2022, <<https://recogito.pelagios.org/>>). See also the resources published on Github, accessed 18 May 2022, <<https://github.com/Travelogues>>.

31 *Travels in the 19th-Century Ottoman Empire: A Digital History Project*, accessed 18 May 2022, <www.middle-east-travelers.de>.

32 Jörg Wettlaufer and Deniz Kilincoglu, *Linked Open Travel Data: Erschließung gesellschaftspolitischer Veränderungen im Osmanischen Reich im 19. Jahrhundert durch ein multimediales Online-Portal*, in Patrick Sahle (ed.), *DHd 2019. Digital Humanities: multimedial & multimodal*.

The methods used in our project are both quantitative and qualitative. While research into the automated recognition of entities such as places and people is based on a quantitative approach, research into the emergence of nationalism is carried out with a mixed approach of quantitative and qualitative methods (e. g., text mining, topic modeling, visualization of entities, and hermeneutic inquiry). The project as a whole, therefore, adopts a mixed-method approach. This approach is becoming increasingly prominent in the field of digital humanities.

For all aspects of the project, the creation of a multilingual historical corpus is a prerequisite. For this purpose, the published reports of European travelers have been identified via Worldcat³³ and other online catalogs, and collected in a Zotero³⁴ bibliography, covering both the primary and secondary literatures. The criterion for the selection of the sources is the existence of a published travelogue of a West European or North American traveler to the Ottoman Empire, which refers to an actual travel experience in the period roughly between 1789 and 1914. For practical reasons, we focus on reports in German, French, and English. So far, we have identified manually about 800 individual reports that fit these criteria. We have completed the preliminary assessments of these reports, especially with respect to the identity of the travelers, and we have built a portal in which the collected and linked information can be queried via a simple user interface. The portal provides a freely accessible bibliography of the reports, collected in a Zotero group library, a working list of travelers with links to Wikipedia entries, outlines of itineraries, sample pictures from printed travelogues, and search functionality across the corpus. In a later stage of the project, a public accessible text-mining environment will be provided. According to initial samples, digitized copies from Google Books and HathiTrust are available for approximately ninety-five percent of the travelogues. Those will be used primarily for the project. So far, a sample of 185 volumes has been downloaded and, if necessary, machine-readable text has been extracted with Optical Character Recognition. Results are satisfactory for our purpose. In particular and rare cases, reports will also be retro-digitized within the project.

Konferenzabstracts. Universitäten zu Mainz und Frankfurt (Frankfurt am Main 2019), pp. 315–317.

³³ Worldcat, accessed 18 May 2022, <<https://www.worldcat.org/>>.

³⁴ Zotero, accessed 18 May 2022, <<https://www.zotero.org/>>.

5 Named entities and text mining

Text mining is a largely automated process of knowledge discovery in textual data, which should enable the effective and efficient use of available text archives.³⁵ The rapid development of the processes of deep learning with neural networks opens up new perspectives in the field of text mining and the recognition of entities. Natural Language Processing (NLP) has changed fundamentally in recent years due to the advances in deep learning and neural networks. Until recently, conditional random fields (CRF) were the standard model for labeling specific word sequences and recognizing named entities. Today, deep learning has replaced CRFs in terms of performance in sequence labeling, and the focus has shifted from feature engineering to the design and implementation of effective neural network architectures. However, just as it is possible and useful to integrate neural network functions into a CRF, CRFs have influenced some of the best deep learning models for labeling sequences. The inclusion of gazetteers, i.e., controlled word lists, can further increase the recognition rate in particular situations and thus further minimize the effort involved in manual post-treatment and correction. Based on word2vec research³⁶ and word or sense embeddings³⁷ with long short-term memory (LSTM) in combination with CRFs, F1 values of approximately ninety percent on modern English-language corpora are possible. Experiments with bidirectional LSTM and convolutional neural networks (CNN) have also shown very promising results.³⁸

One of the challenges in our project is the automated recognition of named entities in a multilingual and historical corpus. The project is therefore based on a combination of different approaches, which overall aims at improved detection and a low error rate.³⁹ The multilingual nature of the corpus is a particular chal-

35 Alexander Mehler and Christian Wolff, *Einleitung: Perspektiven und Positionen des Text-Mining*, in *Zeitschrift für Computerlinguistik und Sprachtechnologie* 20/1 (2005), pp. 1–18; Caroline Sporleder, *Natural Language Processing for Cultural Heritage Domains*, in *Language and Linguistics Compass* 4/9 (2010), pp. 750–768.

36 Tomas Mikolov et al., *Efficient Estimation of Word Representations in Vector Space*, in *Proceedings of the International Conference on Learning Representations (ICLR 2013)*, accessed 18 May 2022, <<http://arxiv.org/abs/1301.3781>>.

37 Jose Camacho-Collados and Mohammad Taher Pilehvar, *From Word to Sense Embeddings: A Survey on Vector Representations of Meaning*, in *Journal of Artificial Intelligence Research* 63 (2018), pp. 743–788.

38 Jason P. C. Chiu and Eric Nichols, *Named Entity Recognition with Bidirectional LSTM-CNNs*, in *Transactions of the Association for Computational Linguistics* 4 (2016), pp. 357–370.

39 We are planning to use frameworks like Apache Spark (accessed 18 May 2022, <<https://spark.apache.org/>>) or TensorFlow (accessed 18 May 2022, <<https://www.tensorflow.org/>>).

lenge. This challenge does not only exist at the work level, but also within texts, for example, if longer passages in other languages are quoted in the text. While NER has already been well researched for modern English with the help of LSTM and CNN, it remains a research desideratum for other languages and, above all, language levels.⁴⁰ Testing NER on historical language levels is also just in its early phase.⁴¹ Another challenge of the corpus is presented by historical place names, which travelers often reproduce through onomatopoeia. These entities must not only be recognized, but also identified in an automated way. The aim here is to research new technologies for the most exact and language-tolerant recognition of entities with minimal post-processing.

In close cooperation between historians, natural language processing specialists, and software engineers, our project aims to explore new ways of increasing recognition rates while creating a generic application profile. The extracted metadata about people, places, publications, and travel routes will ultimately be made available as machine-readable linked open (travel) data. These data will in turn be linked to other resources such as norm data sets like Wikidata⁴² and thematically similar projects mentioned above. Furthermore, the identified entities and travel routes can be displayed using GIS functionality. Incorrectly geolocated data can be corrected easily with the help of draggable markers through crowdsourcing in a graphical user interface.

For the content indexing of the travelogues, the machine-readable texts are processed in an NLP pipeline and relevant entities are extracted. This enables a faceted search in the Apache / Solr index (Fig. 1). The research on the emergence of national identities and sentiments using the corpus of travel reports will be carried out by using and further developing the open source software, Open Semantic Search (OSS). The core of this enterprise search engine is a Lucene SOLR indexer,⁴³ which can be used to index documents of various formats thanks to the integration of Apache Tika.⁴⁴

The information contained in the corpus can thus be found and referenced as Key Word In Context (KWIC). An Open Semantic Extract-Transform-Load (ETL) framework is integrated in the OSS, in which the texts are prepared for extrac-

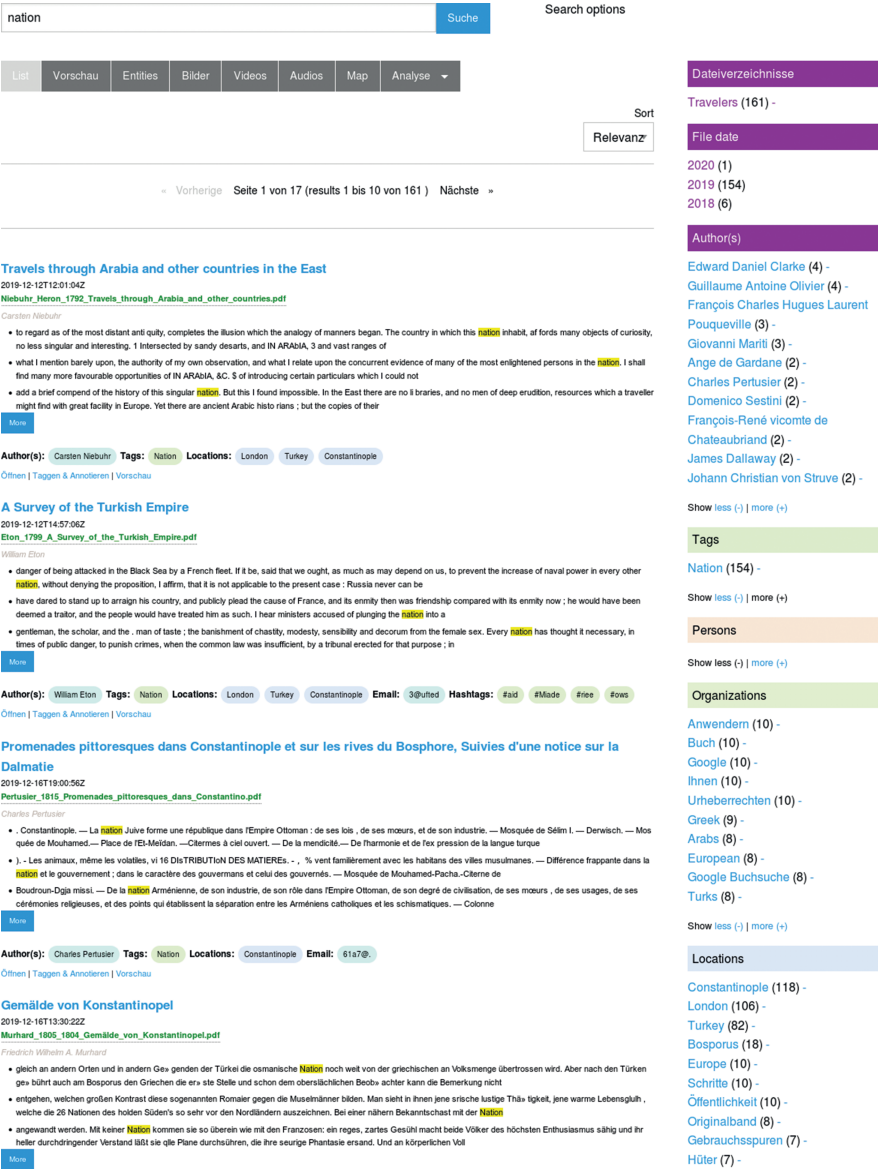
⁴⁰ Sebastian Ruder, Ivan Vulić, and Anders Søgaard, A Survey of Cross-Lingual Word Embedding Models, in *Journal of Artificial Intelligence* 65/1 (2019), pp. 569–631.

⁴¹ Martin Riedl and Sebastian Padó, A Named Entity Recognition Shootout for German, in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, vol. 2: Short Papers (Melbourne 2018), pp. 120–125.

⁴² Wikidata, accessed 18 May 2022, <https://www.wikidata.org/wiki/Wikidata:Main_Page>.

⁴³ Solr, accessed 18 May 2022, <<http://lucene.apache.org/solr/>>.

⁴⁴ Apache Tika, accessed 18 May 2022, <<https://tika.apache.org/>>.



cation as well as a user-friendly interface (Django),⁴⁵ with which the search engine can be operated, configured, and, of course, the data can be searched. The faceted search plays a special role here since the facets are automatically formed from the linguistic analysis of the texts and on the basis of named entities gazetteers (persons, places, nations, etc.) Through the OSS framework, external gazetteers can be fetched from Wikidata and integrated in the ETL process. We assume that the combination of these different methods can increase precision and minimize the effort for post-processing.⁴⁶ The OSS comes with integrated Named Entity Recognition (NER), powered by spaCy or Stanford NER. The out-of-the-box recognition rate with default models is too low to be useful for research purposes. At this point, the NER based on self-trained and better adapted models will be tested in the tool chain and will help to increase precision and recall. More importantly, due to the language-independent method, the developed set of tools can be used in a variety of ways for other corpora.

As an ultimate objective, the platform's adapted functionality will present the extracted and processed texts, images, and metadata in a multimodal form. It will visualize the travelers' itineraries, display the copyright-free graphical materials that are found in these works, or provide links to such material in other projects, such as the Greek travelogues project mentioned above. As a result, the travel experiences will become visible and the texts will be placed in a geographical and medial context.

Conclusions

Travelogues are invaluable sources for historical research. They provide historians with detailed and mostly uncensored first-hand observations on all aspects of life in the time and places that they cover. Besides, they reflect the cultural and intellectual factors that shape the perception of the author, thereby informing us not only of the observed but also of the observer in their historical context. Many historians, therefore, have benefitted from these highly informative and interesting primary sources in their research on both sides of the observation dynamics. However, conventional methods in using travelogues for research purposes have some significant difficulties and pitfalls.

⁴⁵ Django, accessed 18 May 2022, <<https://www.djangoproject.com/>>.

⁴⁶ OSS LOD integration, accessed 18 May 2022, <<https://www.opensemanticsearch.org/doc/datasetmanagement/pendata/>>.

The first difficulty arises from the sheer size of the corpus. Regarding the works on the Middle East, for example, one can find hundreds of travel accounts that cover the decades or the century in question. Considering that each of these works consists of hundreds of pages, we can simply see that one must sift through tens of thousands of pages to find the information relevant to the subject matter. The second difficulty, which exacerbates the first one, is the problem of reliability. Many travelogues contain information and at times even entire passages copied from earlier works, thereby not reflecting real observations. Some authors add fantastical elements in their accounts to create sensation, mainly for marketing purposes.⁴⁷ In some cases, authors wrote entirely fictional accounts that pose as travelogues.⁴⁸ All these misleading narratives create serious pitfalls for the researchers who would like to benefit from these otherwise unrivaled historical sources.

Although it is not easy to avoid all such risks entirely, we believe that digital tools and methods provide us with considerable potential to minimize them. In this respect, we need to develop strong yet intuitive tools to open these vast and mostly untapped historical sources to the use of researchers. User-friendly tools and interfaces for simple and complex search queries, intertextual analyses, and especially for data extraction, organization, and visualization purposes will expand the horizons of historical research in all disciplinary subfields.

With these broad objectives in mind, we set out to develop a digital history project on nineteenth-century travelogues about the Middle East, and particularly about the Ottoman Empire. Following a mixed-method approach, which includes both digital and conventional historical methods and questions, we aim to identify and compare the descriptions of the sociopolitical conditions and developments of the age, observed by foreign travelers. Our project focuses particularly on the rise of nations and nationalisms in this era. However, it aims at building a digital research environment for researchers who work on all aspects of life in the nineteenth-century Middle East.

Although we are still in the early phases of our project, and it still needs considerable amount of time and funding to reach its goals, we have made significant observations regarding the potential of digital research on travelogues. We

⁴⁷ For a critical analysis of such a nineteenth-century travelogue on Egypt, which “transcended the established boundaries of the travel genre in order to become part of nineteenth-century popular entertainment culture,” see Daniela Richter, *Inside the Oriental Spectacle: Hermann von Pückler-Muskau’s Egyptian Travelogue*, *Colloquia Germanica* 46 (2013), pp. 229–244.

⁴⁸ For one of the most “successful” medieval examples, namely the *Book of Mandeville*, see Sharon Kinoshita, *Medieval Travel Writing* (2): *Beyond the Pilgrimage*, in Das and Youngs (eds.), *Cambridge History of Travel Writing*, pp. 59–61.

extracted some illuminating observations on (proto)nationalist tensions and sentiments, for example, using digital, but still manual, methods. These examples testify to the unrivaled and hitherto untapped wealth of these sources, which can be extracted fully and organized effectively only through digital methods. Based on these observations, we plan to develop a multilingual digital corpus and the necessary automatic and semi-automatic digital tools to extract metadata about people, places, publications, and travel routes from this corpus. Our goal is to link these metadata to each other and also to other similar resources and develop more advanced tools to make this knowledge base available as machine-readable linked open (travel) data. In addition, we plan to develop a portal with intuitive tools for data extraction and visualization to offer an open, rich, and reliable resource for all scholars and research groups interested in the nineteenth-century Middle East. Finally, as another ultimate goal, we envision to reuse the specific adaptations in the NLP Toolchain of the OSS search environment for prospective research projects that use all sorts of historical texts from this period.

