



UNIVERSITY OF OXFORD

**Oxford Economic and Social
History Working Papers**

Number 201, October 2022

**African Time Travellers: What can we learn from 500
years of written accounts?**

Edward Kerby, Alexander Moradi, and Hanjo Odendaal

African Time Travellers: What can we learn from 500 years of written accounts?

Edward Kerby*

Alexander Moradi[†]

Hanjo Odendaal[‡]

October 2022

Abstract

In this paper we study 500 years of African economic history using traveller accounts. We systematically collected 2,464 unique documents, of which 855 pass language and rigorous data quality requirements. Our final corpus of texts contains more than 230,000 pages. Analysing such a corpus is an insurmountable task for traditional historians and would probably take a lifetime's work. Applying modern day computational linguistic techniques such as a structural topic model approach (STM) in combination with domain knowledge of African economic history, we analyse how first hand accounts (topics) evolve across space, time and traveller occupations. Apart from obvious accounts of climate, geography and zoology, we find topics around imperialism, diplomacy, conflict, trade/commerce, health/medicine, evangelization and many more topics of interest to scholarship. We find that some topics follow notable epochs defined by underlying relevance and that travellers' occupational backgrounds influence the narratives in their writing. Many topics exhibit good temporal and spatial coverage, and a large variation in occupational backgrounds adding different perspectives to a topic. This makes the large body of written accounts a promising source to systemically shed new light on some of Africa's precolonial past.

Keywords: Africa, Economic History, Explorer, Travelogue, Text Analysis, Structural Topic Modelling

JEL Codes: C55, N01, N17, N37, N57, Z10

1 Motivation and Historiography

In the recent two decades, there has been a surge in studies in the field of African Economic History.¹ Many studies focus on the colonial period when bureaucratic apparatuses became developed enough to produce government documents. The precolonial period, in contrast, has remained underrepresented due to the dearth of sources that scholars could engage with. The anthropological observations summarized in [Murdock \(1959\)](#)'s *Ethnographic Atlas* have remained the main data source for quantitative studies of precolonial Africa, even though observations are overwhelmingly drawn from the colonial period.²

*Department of Economics, Stellenbosch University & Laboratory for the Economics of Africa's Past (LEAP), South Africa. Corresponding author - EdKerby@sun.ac.za

[†]Department of Economics, University of Bozen-Bolzano, Italy. Previously affiliated with Nuffield College and Oxford Centre for Economic and Social History.

[‡]Department of Economics, Stellenbosch University, South Africa.

Acknowledgments: Thanks to Johan Fourie, Max Roser, Elliott Ash, Sascha Becker, and participants from the 3rd Monash-Warwick-Zurich Text-as-Data Workshop, and the Stellenbosch Development Data Visualisation Conference. We thank Tiaan De Swardt and Cailin McRae for research assistance in coding and collecting travel diaries across Africa. We also thank the University of Bozen-Bolzano for funding, and the Laboratory for the Economics of Africa's Past (LEAP) at Stellenbosch University, South Africa. Visit the project at www.timetraveller.voyage.

¹For overviews see [Austin \(2016\)](#); [Moradi \(2018\)](#); [Michalopoulos and Papaioannou \(2020\)](#).

²These data were subject to criticism (e.g. [Kopytoff, 1988](#)). [Bahrami-Rad et al. \(2021\)](#) showed that there is a strong correlation between Murdock's ethnographic data and contemporary self-reports in survey data - though this finding is not very useful regarding the accuracy of Murdock's data reflecting *precolonial* institutions. Only 7.5% of Murdock's observations for Africa were made before 1900.

In this paper, we explore a source that can potentially shed light on 500 years of Africa’s precolonial past: travel diaries. The value of this source has been recognized since the 1980s (Jones, 1987; Bridges, 1987). The historical travel literature is immense. Travellers tended to write about a large number of topics ranging from land, people and places, to weather observations. This collective knowledge of multiple journeys and travellers can bring diverse views and observations.³ In spite of this, the source has never been systematically used for the whole of the African continent.⁴ This is owed to the copious volumes of travel descriptions, and richness in text. Jones (1987) speculated that it may be possible with “the help of modern technology”, to aggregate multiple sources and represent the results “in detail on a set of maps”. In recent years, these travel diaries became accessible online through libraries and websites like archive.org. Indeed, as Jones (1987) foresaw, we now have the computing power and statistical techniques to analyse thousands of traveller accounts in order to gain some first hand insight into the economic history of Africa over the last half millennium. Figure 1 shows an indication of the potential spatio-temporal coverage.

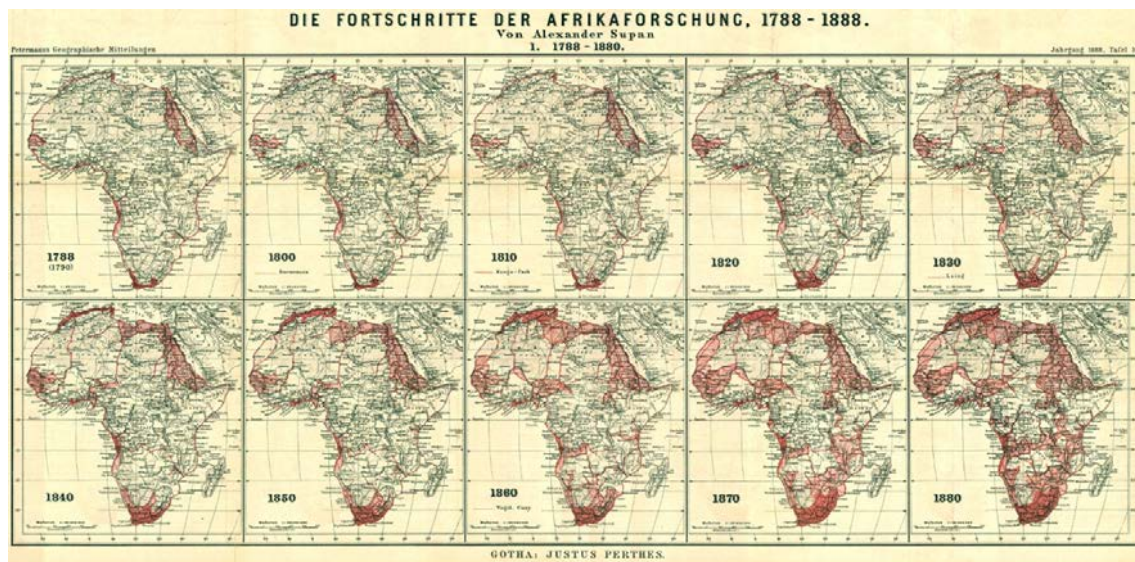


Figure 1: Supan (1888) showing the expanding knowledge of Africa in the Western world.

Of course, there are potentially many drawbacks. Information comes selected and filtered through the eyes of the traveller.⁵ The diaries were not necessarily aimed for the scientific community but to entertain the curiosity of a paying readership or the traveller’s sponsor.⁶ Descriptions may be subjective. There

³Jones (1987, p. 65) notes that available single explorer sources can be “one-sided [...] dealing exclusively with slavery or [...] even merely with flora and fauna.”

⁴There are biographies of the more famous explorers such as Livingstone and Stanley (e.g. Ross, 2002; Driver, 1991; Blunt, 1994; Jeal, 2007; Berenson, 2011). Then there are numerous works that shed light on subsets of countries and explorers (Hibbert, 1982; Haarhoff, 1989; Rotberg, 1970). A big literature deals with the age of exploration more generally (Curtin, 1984; Fabian, 2000). Finally, philologists are interested in the travelogue as a genre (Haarhoff, 1989; Stiebel, 2002; Murray, 2016; Youngs, 1994).

⁵One concern is certainly that these were mostly European observations. However, we do not believe that this justifies disregarding the source entirely. First, Africans were not bystanders in travel. Europeans enlisted the help of Africans and relied on information Africans provided; often following established trade routes (Jones and Voigt, 2012; Rockel, 2006) with African porters and guides sometimes the veterans of multiple expeditions (Shackleton et al., 1972; Simpson, 1975; Bridges, 1982). For this reason, we also use the term “travellers“ and avoid the notion of “explorers”.

⁶Hallett (1963) notes that behind most travellers stood a sponsor, prepared to invest capital in the traveller’s enterprise.

may even be intentional misrepresentation. Observations and the choice of topics may have followed fashions in Europe. Hence, the journey's dangers may be exaggerated, facts distorted to an extent that diaries resemble more fiction than reality (Spurr, 1993). Bassett and Porter (1991) note that similar problems are associated with explorers' attempts at cartography.⁷

Nevertheless, one should not dismiss the wealth of information recorded by travellers. There are a number of possible applications in quantitative studies. Nunn and Wantchekon (2011) used the routes of 25 famous explorers to measure precolonial contact with Europeans. Nicholson et al. (2012) used meteorological observations from explorers to reconstruct rainfall over two centuries. Linares (2002) reconstructed the extent and practices of rice cultivation in West Africa around 1500. Irwin (1975), Smith (1973) and Bridges (1963, 1973, 1982, 1987) used traveller diaries to shed light on precolonial African diplomacy; they still remain seminal contributions today.

A database of traveller accounts may also have exceptional promise for new applications. Using traveller's routes, one may reconstruct a more complete and detailed precolonial trade route network including modes of transport, travelling time, disruptions due to conflict or trade shocks over time. Observations on the fauna and flora may shed light on environmental changes (Pooley, 2016). Slavery and the Trans-African slave trade may be quantified at a more local level (Whatley, 2021). By extracting places that travellers visited, it may be possible to expand the Chandler (1987) database of cities in Africa. Finally, we may cross-check the Murdock (1959) Atlas and test claims about persistence of ethnic institutions (Bahrami-Rad et al., 2021) or disruptions caused by colonization (Jerven, 2011; Henderson and Whatley, 2014).⁸

The interest in travel diaries as a source declined in the 1990s. The aim of this article is to revive interest in this source and establish what we can learn from traveller accounts. Firstly, we build a database of travellers and their written accounts of Africa using rigorous data quality checks. After an extensive cleaning process, we apply a structural topic model approach (STM) to the corpus in order to gain insight into how written themes (or topics) have evolved across space, time and traveller occupation. Apart from obvious accounts of climate, geography and zoology, we find topics around imperialism, diplomacy, conflict, trade/commerce, health/medicine, evangelization and many more topics of interest to scholarship. Some of those topics are covered by multiple travel accounts and span centuries. Unsurprisingly, we find that the background of travellers matters for the choice of topics. Military men write more about conflict and diplomacy; missionaries more about Christian values and evangelization; civil servants more about imperialism and European settlement. This diversity is a strength of the source. Fortunately, while the travellers' background undergoes cycles over time, it appears that there is

Up to the middle of the 18th century the sponsorship of exploration had come either from the State, Church or a commercial organization.

⁷The standard example for significant errors is the depiction by Park and Rennell (1799) of the Mountains of Kong, which never existed but were reproduced in 19th century maps, again and again.

⁸Indeed, Thornton (1983) argued that anthropology has its roots in the 19th century travelogue and missionary letters.

always a critical mass of the most important occupations available for temporal comparisons to be made. Overall, we argue that the large body of written accounts can be used to shed new light on Africa’s past. Ultimately, the accounts are so numerous, that they may provide for a database that can grow and be flexibly used for a wide range of enquiries. Furthermore, there may be potential to apply the analysis to other regions and times.

2 Data

A directory of travellers to Africa does not exist. Hence, our first task is to build one. We define travellers as individuals who write about their journeys. These include state and non-state actors. We restrict our study to the precolonial period and exclude journeys after 1900.⁹

2.1 Building a Database of “Travellers of Africa”

Step One of the process entails searching for items available at three large repositories for online historical books and archives, namely [Internet Archive](#)¹⁰, [Gallica](#)¹¹, and [Google Books](#).¹² We used the following combination of search terms:

(voyage* OR exploration* OR expedition* OR travel* OR survey* OR excursion*) AND
LOCATION AND (date: 1200 TO 1900 OR NA) AND (mediatype = texts)

where **LOCATION** is defined from a compiled list of historical names of African regions and territories (see [Appendix A](#)).¹³

Step one added 1,971 travellers, 1,534 documents and 1,476 journeys to our database.

Step Two of the process entails adding travellers that were not identified in *step one*, e.g. because i) the search terms were too broad ii) travellers did not leave a diary or iii) the diary has not yet been digitized. We manually searched several secondary sources.¹⁴ We cross referenced these to the meta search to exclude duplication. This added 257 travellers, and 198 additional documents to our database.

⁹European powers formalised the colonisation of Africa during the Berlin Conference 1880-1885. By then large parts of Africa were still not under effective colonial control. Hence, we chose 1900 rather than 1885 as the end of our study period.

¹⁰The Internet Archive is one of the world’s largest public digital libraries, with an extensive collection of human culture: 2 million books.

¹¹Gallica is one of the major digital libraries available for free via the Internet.

¹²The Google Books Library Project is aimed at scanning and making searchable the collections of several major research libraries.

¹³Time stamp for the most recent meta search across the online archives was 17/11/2021.

¹⁴Secondary sources include; 1) Botanical Exploration Southern Africa ([Gunn and Codd, 1981](#)), 2) Political Power in Precolonial Buganda ([Reid, 2002](#)), 3) Dictionary of National Biographies ([Matthew et al., 2004](#)), 4) Pre-Colonial Africa in Colonial African Narratives ([Wehrs, 2008](#)), 5) The London Geographical Journal ([Royal Geographical Society of London, 1839](#)), 6) The Royal Geographical Society([Royal Geographical Society, 1844](#)), 7) German men in Africa ([Weidmann, 1894](#)), 8) Mapping the History of cotton textile production in precolonial West Africa ([Kriger, 2005](#)), 9) Transactions of the Royal Society of Literature ([Royal Society of Literature, 1827](#)).

Step Three, we ‘hand coded’ metadata variables from each document, with the following used in this paper: Traveller’s occupation, and country of nationality as first listed in Wikipedia or in the original source. Region of travel defined by the journey entry point, as interpreted from the original source and accompanying maps.¹⁵ Publication date of the original document. Original language as printed in the source, and if an English language translation is incorporated to the final corpus.

2.2 Background of Travellers

Using our database, we describe the travellers’ background and how it changed over time. Their occupations were diverse. We recorded 199 unique occupational titles that we group into 10 major categories: scientist, military, missionary, civil servant, medical, explorer, writer, merchant, aristocrat, and other.

Table 1 provides an overview of the number of documents and journeys by traveller’s occupation listing the three most frequent job titles in those categories. Among the most frequent occupations are scientists (16.1%), military men (14.8%), missionaries (10.6%), civil servants (8.8%) and writers (8.8%). The term “explorer”, associated with 5.7% of the documents, may be a generic description of early travellers. Merchants (3.4%) and aristocrats (2.2%) are also present.

Table 1: Traveller occupations in the corpus.

Occupation Cat.	Most common titles	Documents	Journey	Diff.
Explorer	Explorer	5.70%	6.10%	-0.40%
Military	Captain / Army Officer / Soldier	14.80%	17.51%	-2.71%
Merchant	Merchant / Slaver / Trader	3.39%	3.22%	0.17%
Missionary	Missionary / Reverend / Clergyman	10.56%	13.38%	-2.82%
Scientist	Botanist / Naturalist / Archaeologist	16.12%	17.64%	-1.52%
Civil Servants	Colonial Admin. / Diplomat / Politician	8.81%	7.99%	0.82%
Writers	Journalist / Traveller / Travel Writer	8.81%	7.99%	0.82%
Aristocrats	Duke / Aristocrat / Esquire	2.21%	1.71%	0.51%
Unknown	NA	16.65%	3.62%	-13.02%

We then calculate the cumulative proportion over time. The cumulative proportion can be interpreted as the acquired journeys by a group at a specific point in time.¹⁶ Conceptually, it is closer to acquired knowledge at time t . Moreover, the measure is less noisy than occupational frequencies by half-centuries and deviations from the norm can be conveniently spotted at the decade they occurred.

Figure 2a shows that three occupational classes dominated the early journeys: military men, merchants and explorers. The prevalence of the latter two steadily declined over time. Military men, in contrast, remained prominent. Indeed most travellers to Africa had a military background; their proportion remained relatively stable from 1670 until 1840 when their frequency fell from 20% to 15%.

¹⁵We assigned a globally standardised region meta variable to journeys as per Appendix A.

¹⁶The cumulative proportion at time t is calculated as the cumulative frequency of journeys X by group i divided by the total frequency of journeys by time t : $Prop_{i,t} = \sum_{t=0}^{1900} \frac{X_{i,t}}{N_t}$.

Missionaries were frequent between 1650 and 1770, declined then only to bounce back around 1840. The date coincides with the introduction of quinine in travel chests and the intensification in evangelization efforts in Africa (Jedwab et al., 2022; Curtin, 1998). Interestingly, missionaries appear to have mostly replaced men of military background. Finally, the share of scientists increased around 1700, gained momentum around 1760, and remained high at about 15%.

Figure 2b describes the travellers’ nationality. Despite our English language restrictions, the nationality of travellers very much reflects the history of imperial endeavour. The Portuguese were at the vanguard of the age of exploration. The Dutch increased their involvement during the Dutch Golden Age (1588-1672), head-to-head with the British until 1730, after which their influence collapsed. The British and French influence dominates during the scramble, with the Germans as latecomers, only making up a small share.

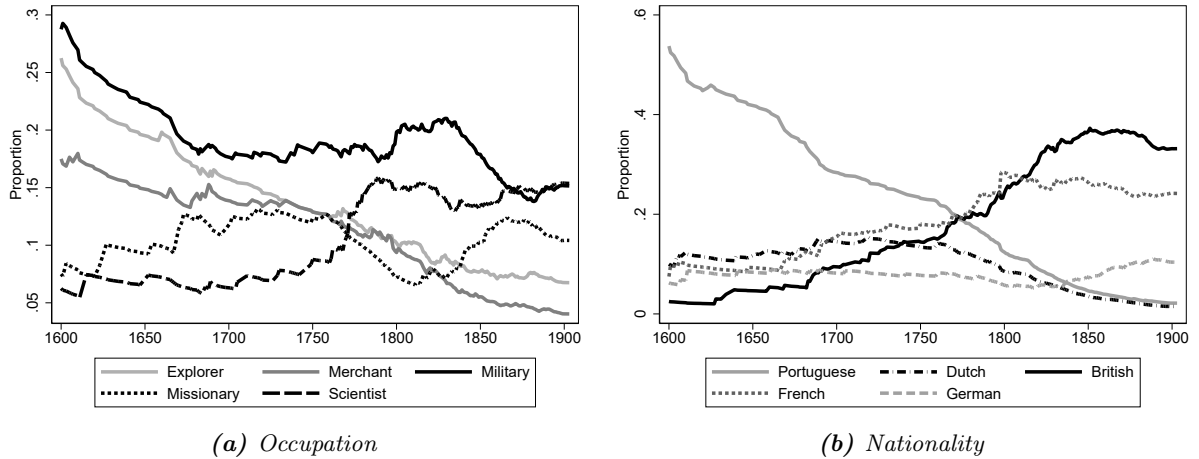


Figure 2: Traveller’s Background over Time.

3 Structural Topic Modeling

Traditionally historians were burdened with reading vast amounts of historical records in order to get a quantitative synopsis of culture, technological progress or ethnographic evolution. In recent years, with the increase of computing power and availability of machine-readable text, this burden has been shifted to machines in what has evolved into a field called computational linguistics.¹⁷ The application of these techniques have become especially prominent within the analysis of recorded ideas and social accounts (see e.g. Blaydes et al. (2018); Lucas et al. (2015); Roberts et al. (2014) and Grajzl and Murrell (2019)). In this paper we analyse 500 years of written accounts using topic modeling, a well known natural language processing technique within computational linguistics. The key characteristic of

¹⁷Fourie (2016) summarises recent analytical advances in economic history but concludes that “in the history profession, this data revolution has gained only limited traction.”

this technique is that it captures latent clustering of tokens, or “topics” using an unsupervised machine learning algorithm and a corpus of compiled text. The main argument for using topic models is to aggregate core ideas from a large corpus of text, which would be too labour intensive for historians wishing to better understand changing social constructs over centuries of written accounts.

There are different approaches to topic modelling. We follow an approach first introduced by [Roberts et al. \(2014\)](#) called Structural Topic Modeling (STM). The *structural* component of the model relates to the fact that, unlike traditional topic modeling that uses Latent Dirichlet Allocation (LDA), meta information is brought into the generative process and employs correlated topic models (CTM).¹⁸ The addition that [Roberts et al. \(2014\)](#) made to the traditional topic modeling approach is to dictate that the topic-prevalence vector should be a function of additional information as opposed to the earlier versions where it was drawn from the same distribution for all documents. Incorporating metadata into the model allows the researcher to estimate the relationship between document characteristics (metadata) and topic prevalence. In most cases, the relationship between the metadata and the latent topics are in essence the estimand of scientific interest. To put this plainly, structural topic models not only help us to understand what is *within* the text, but also helps us leverage information *about* the text. If we think about the mechanics, a “structural” topic model aims to build topics using a probabilistic distribution over tokens and their co-occurrences with one another conditioned on various metadata associated with a given document.

Mathematically this can be represented by denoting a collection of documents as D , indexed by d , with each of the documents containing w_d words. The unit of analysis within this paper is a single digitized *page*. A topic k can then be defined by the proportion of distinct words (or tokens) V (equation 1), which is known as the topic prevalence matrix (θ). The number of topics K is a subjective choice by the analyst.¹⁹ Each of these topics are associated with a V -dimensional probability mass function, β_k (topic content matrix), that controls the frequency according to which tokens are generated from that topic (equation 2).

$$\theta = \begin{array}{c|cccc} & Topic1 & Topic2 & \dots & TopicK \\ \hline Doc1 & 0.20 & 0.01 & \dots & 0.01 \\ Doc2 & 0.20 & 0.03 & \dots & 0.02 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ DocD & 0.02 & 0.20 & \dots & 0.04 \end{array} \quad (1)$$

¹⁸See [Blei and Lafferty \(2006\)](#) for overview of how CTMs extend LDA.

¹⁹ Although some progress has been made to computationally decide on the number of K , there is no way of estimating a *best* K ([Arun et al., 2010](#); [Grimmer and Stewart, 2013](#)). See Appendix E for a visual illustration of where we used a data-driven approach to decide the final K . Among the test performed were held out likelihood estimation by [Wallach et al. \(2009\)](#), as well as [Taddy \(2012\)](#)’s residual analysis.

$$\beta^{\mathbf{T}} = \begin{array}{c|cccc} & \textit{Topic1} & \textit{Topic2} & \dots & \textit{TopicK} \\ \hline \textit{text} & 0.020 & 0.010 & \dots & 0.001 \\ \textit{data} & 0.220 & 0.030 & \dots & 0.002 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \textit{analysis} & 0.020 & 0.204 & \dots & 0.004 \end{array} \quad (2)$$

where the $D \times K$ matrix in equation (1) defines the topic prevalence, and $V \times K$ matrix in equation (2) defines topical content. The generating process of a single document d , from K topics and observed vocabulary V with words $w_{d,n}$ of a STM model can be summarised as (Roberts et al., 2014):

1. Draw the document-level attention to each topic from a logistic-normal generalized linear model based on a vector of document covariates design matrix \mathbf{X}_d and K -dimensional hyper-parameter vector σ using the topic prevalence model:

$$\gamma_k \sim \text{LogisticNormal}_P(0, \sigma_k^2 I_p), \quad (3)$$

$$\theta_d \sim \text{LogisticNormal}_{K-1}(\Gamma' X'_d, \Sigma) \quad (4)$$

where the document covariate \mathbf{X}_d is a $1 \times p$ vector, $\Gamma = [\gamma_1 | \dots | \gamma_k]$ is a $P \times K - 1$ matrix of coefficients for topic prevalence model and Σ is a $K - 1 \times K - 1$ covariance matrix, while the core language model is constructed as:

$$z_{d,n} \sim \text{Multinomial}_{K-1}(\theta_d), \quad (5)$$

$$w_{d,n} \sim \text{Multinomial}_V(\mathbf{B}_{d,n}) \quad (6)$$

As can be seen from equation (3) to (6), the model allows for correlations in the topic proportions using the Logistic Normal distribution (Aitchison, 1982). This means that for the topic proportion vector, θ_d , for each word within document d , a topic is sampled from a multinomial distribution $z_{d,n} \sim \text{Multinomial}(\theta_d)$ and conditional on a topic, a token is chosen from the appropriate distribution over terms $\beta_{z_{d,n}}$ represented by $\mathbf{B}_{d,n}$.

The difference between Blei and Lafferty (2007) and Roberts et al. (2014), is that in the latter, the μ_d (prevalence) and $\beta_{z_{d,n}}$ (topical content) are no longer global parameters shared by all documents, but are optional inputs specified as document-level covariates.

Modeling how the topics are affected by the covariates, the word occurrences are measured in terms

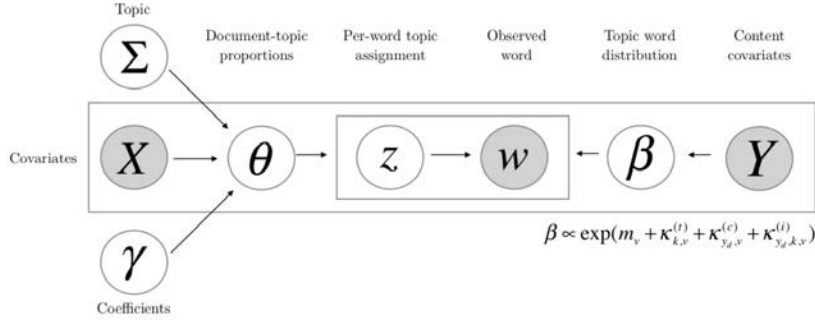


Figure 3: Plate diagram of a structural topic model, incorporating topic prevalence and topic content. Source: Roberts et al. (2014)

of the log-transformed rate deviations from the corpus-wide background distribution m . In this model, the log-transformed rate deviations are denoted by a collection of parameter’s, represented by κ , where the superscript indicates the set they belong to: t for topic, c for covariates or i for topic-covariate interactions.

3.1 A Corpus of Traveller Documents

Once we formalized the method for the collection of documents, we built a machine based production pipeline through which the documents are ‘digested’ in order to be digitized and archived within a structured database. Table 2 provides an overview of the steps in building the corpus and the corresponding volume of journeys at each stage of the automated pipeline. We found 2,171 traveller accounts and uploaded the PDF’s into the analytical pipeline. Of these, 854 passed the OCR English language test representing 237,741 pages of text. To prepare the data for analysis, commonly used cleaning procedures such as removing punctuation, digits and stop-words was applied.²⁰ Throughout the paper we also refer to words as tokens, as the concept generalizes more broadly. A token can either be composed of one word, two words (i.e. “explorer accounts”) or larger terms. These are commonly known as N -gram tokens, where N represents the number of tokens. The full list of pre-processing steps are described in detail in Appendix B, and read together with Appendix C provides a visual overview of the process flow constructed to estimate the STM presented in Section 3.

Table 2: Building the Corpus

	Description	Documents
Step 1	Directory of Travellers	2,171
Step 2 + 3	Library of text from search process	2,464
Step 4	OCR texts in English	1,298
Step 5	Pass all data quality checks	855
Working Corpus:	237,741 pages	Median: 117 tokens/page

²⁰Stop-words constitute words like “and”, “the”, “a”, “that”, “am” that by themselves are not informative and only introduce noise into the modeling process. We used the SMART word list from Lewis et al. (2004) as stop-word dictionary.

Table 3 contains summary statistics of the working corpus finally used in the topic model. The coverage naturally reflects the different timing and extent of the underlying Africa journeys. The number of pages per journal and tokens per page is similar across regions, the average being 278 pages and 117 tokens per page.

Table 3: Descriptive statistics of working corpus

Region	Years (Min)	Years (Max)	Years (Avg)	Documents (#)	Pages (#)	Tokens (#)	<tokens/page>
Central	1578	1894	1867	41	11,039	1,407,122	127
East	1327	1899	1843	141	39,540	4,975,164	126
North	1396	1899	1837	381	109,306	13,649,799	125
South	1593	1899	1847	153	40,639	5,457,903	134
West	1583	1899	1834	139	37,217	4,650,126	125
Totals				855	237,741	12,859,511	117

To illustrate the spatio-temporal component of the corpus, Figure 4 highlights the cumulative collection of documents in the corpus. It closely resembles the map of knowledge of Africa in the western world drawn by Supan (1888) and displayed in Figure 1.

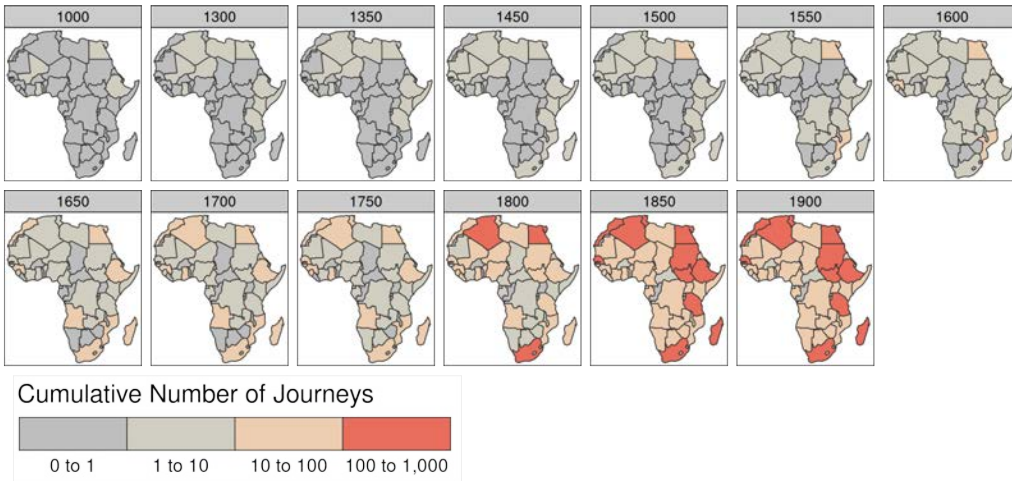


Figure 4: Cumulative journeys throughout Africa using our collected corpus (1000 AD - 1900).

3.2 Publication bias

Only a subset of journeys entered our STM model. Does this create a bias? We investigate this using a probit model.

$$Pr(Y_i = 1|X) = \Phi(X\beta + e_i) \quad (7)$$

where the dependent variable, Y_i is a dummy indicating whether document i entered our STM

analysis. The vector of covariates X consists of four groups of variables measuring the journey’s date and region as well as the traveller’s nationality and occupational background. All the covariates are dummies. To facilitate interpretation we report marginal effects. Standard errors are clustered at the traveller level. First, certain travellers attracted more interest and most or all of their accounts became translated and widely available.²¹ Second, certain covariates (the traveller’s nationality and occupation) do not vary across journeys and documents.

Table 4 reports chi-square tests of joint significance. The first four columns test for (1) Nationality (French, German, Other), (2) Period (1400-1500, 1501-1800, 1801-1900), (3) Region (West, East, Central, South) and (4) Occupational classifications (scientists, military, civil servants, missionary, medical, merchant and other). Considered in isolation, our STM corpus appears to deviate from our traveller database from step 1 in all of the four dimensions. However, the full model in column (5) containing all of the variables reveals that deviations are due to nationality and date that - if ignored - create distortions in the regional and occupational composition.

Table 4: Traveller Documents entering STM

	(1)	(2)	(3)	(4)	(5)
1. Nationality (French, German, Other)	62.91*** (0.000)				425.6*** (0.000)
2. Period (1400-1500, 1501-1800, 1801-1900)		36.69*** (0.000)			54.82*** (0.000)
3. Region (West, East, Central, South)			21.57*** (0.000)		1.17 (0.884)
4. Occupation (sci., mil., civ., miss., med., merch., oth.)				42.39*** (0.000)	8.01 (0.332)
Observations	2,392	2,392	2,392	2,392	2,392
Pseudo R2	0.169	0.170	0.014	0.022	0.513

Notes: This Table reports the results of a probit regression with a dummy variable indicating the inclusion of a document in the STM analysis as dependent variable. Coefficients are chi-square tests of joint significance. “Nationality“ consists of dummy variables for French, German and Other. “Period“ are dummies for the Middle Ages (1400-1500), Age of Enlightenment (1501-1800) and Age of Imperialism (1801-1900). “Region“ are West, East, Central, South. “Occupation“ are scientists, military, civil servants, missionary, medical, merchant and other. p-values in parentheses.

*** p<0.01, ** p<0.05, * p<0.1.

Figure 5 plots the marginal effects from column (5) of Table 4. Travellers of non-British origin are underrepresented. While 84.9% of the British traveller’s accounts end up in our STM, the percentage of other nationals is significantly lower (French: -44.9%, German: -39.9%, Others: -16.7%). This certainly reflects our English language requirement. It is worth noting that translations are time consuming and costly, with traveller societies such as Hakluyt often translating the most exciting, famous, patriotic or first of the explorers (Sherman, 2004; Bridges, 1987).²² In addition, early accounts from before 1400

²¹Indeed, standard errors decrease substantially after clustering.

²²The **Hakluyt Society** published scholarly editions of primary records of voyages and travels, widely prized for their standards of scholarship and book production. (Sherman, 2004, p. 199) notes the “role played by these translations [...] is crucial to understanding the intellectual outlook of those involved in overseas venture.”

disproportionately dropped from our analysis. These were possibly in Arabic and saved in non-readable picture format. We believe that advantages of data science methods appear - as of yet - less striking for the early period because the corpus of available texts is so small that traditional approaches are well-suited for analyses. In contrast, there is no significantly different attrition for 1400-1500, 1501-1800 and 1801-1900 periods. In sum, we conclude that biases with respect to occupations, regions and even time do not appear substantial.

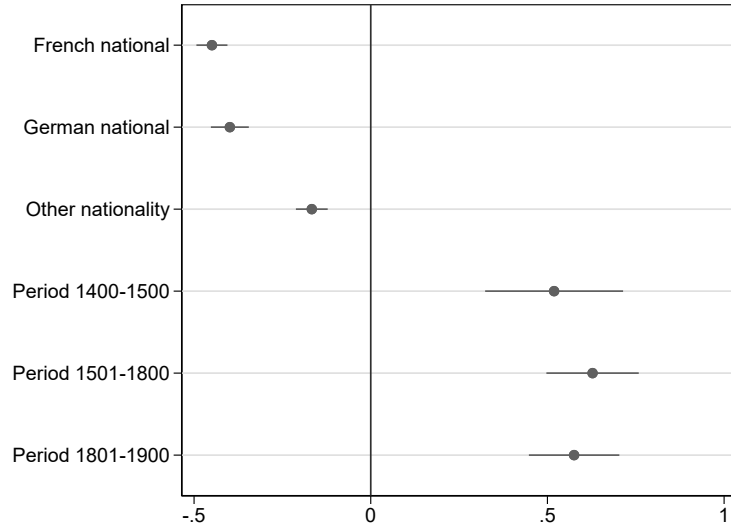


Figure 5: Average marginal effects

Notes: This Figure shows average marginal effects and 95% confidence intervals for selected variables from column (5) of Table 4. Reference category is British, Date 0-1400, North, Explorer.
 *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

3.3 The Metadata

The final step in creating our corpus is integrating the data-driven machine process of the STM described in Section 3 with the travellers’ metadata (i.e. spatial, temporal, occupational, language and nationality) described in *Step Three* of Section 2.1. The integration of metadata covariates to each document allows for a more complete model, which in turn facilitates examining the effect that each of the covariates have on topic prevalence in the final output. More specifically, we allow the topic prevalence to be influenced by our covariates — *traveller occupation*, *time period* and *region* — and as such estimate a geographical, time-variant topic model that also takes into account the occupation of the person writing the text. By introducing this structure into the generative model, we wish to control (and test) for the observational bias first reported by Jones (1987), while also investigating how traveller accounts differed across space and time. The final output of the model will thus allow a varying prevalence of each topic (element of θ_d) for each document, dependent on the linear combination of our dummy covariates as specified in equation (3).

4 Representation of topics in the STM

STM creates the topics, but does not tell us how to label/categorise them. The analyst does it. For this, historical domain knowledge of the underlying body of topics is needed. Moreover, there is no absolute measure of determining the correct number of unique topics for the STM. Given the results of the statistical test, we decided to provide analysis of 25 topics.²³

We now turn to the machine generated output from the STM. Figure 6 summarises the average proportional topic prevalence ($\tilde{\theta}$) across our corpus as well as the ten words with the highest probability of being associated with our 25 uniquely assigned *topic labels*. For example, what proportion of the total text can be attributed to writing that deals with themes around imperialism? STM would suggest it was 6.06%. Appendix F shows a more detailed word list for the 25 topics using four topic word filters (Roberts et al., 2014):²⁴

- (1) *Highest Probability* words are those that are most common for a given topic, but are also nonexclusive, in the sense that they may be associated with any number of topics and might be the highest probability words in more than one topic (as is the case, for example, with ‘king’, ‘law’, ‘man’, ‘will’ in our corpus).
- (2) *Lift* weights words by dividing their frequency in other topics, giving higher weight to words that appear less frequently in other topics.
- (3) *Score* measures the top words, dividing the log frequency of the word in the topic by the log frequency of the word in other topics.
- (4) *FREX* reports words for a given topic that reflect a weighted combination of two criteria: association with that topic with a high probability (frequency) and a greater likelihood of being associated with that topic than with other topics (exclusivity).

All four of these statistical measures for topic word prevalence are useful. In our discussion, however, we mostly use Highest Probability and FREX for interpretation.

In the following section 4.1 we discuss the primary output of the STM: what travellers wrote about, and the concentration of topics across space and time. In section 4.2 we analyse the interconnectedness of topics, while in section 4.3 we use an econometric framework to investigate how the metadata variables (occupation, epoch and region) affect the content generation process.

²³To inform on the “correct” number of topics (K) to investigate, we employ commonly used statistical tests. See Appendix E for a graphical depiction, or Footnote 19 where it is described in detail.

²⁴Appendix F gives the statistical methods for the weighting schema.

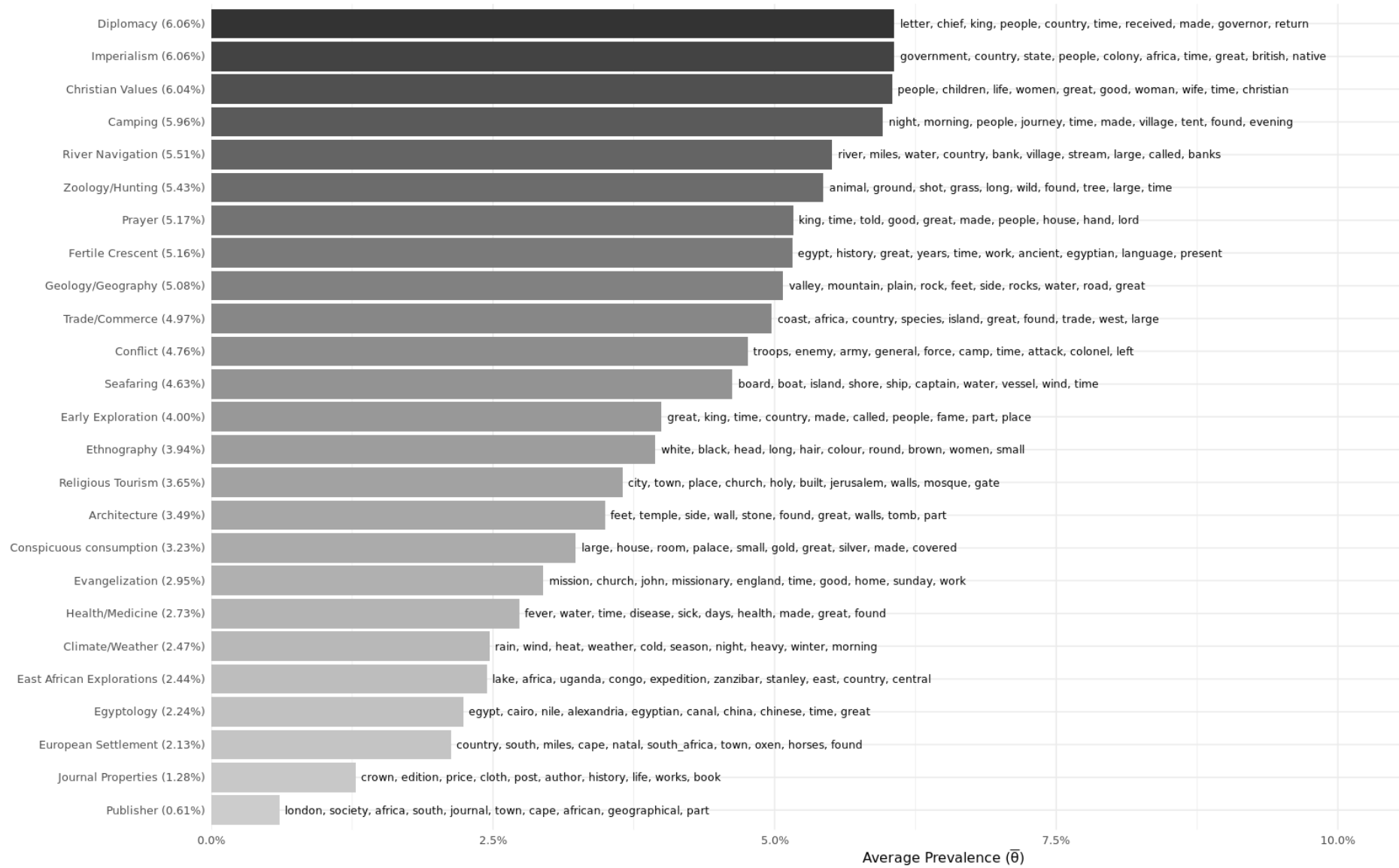


Figure 6: Average topic prevalence ($\bar{\theta}$) across the whole corpus.

4.1 What did travellers write about? When? And where?

Bringing historical knowledge to bear we now examine the topics that STM identified in the corpus of documents.

Imperialism, Diplomacy, Conflict & European Settlement: We define imperialism as a state policy or practice of extending power, by direct territorial acquisition or by gaining political and economic control over another country or territory (based on [Britannica](#)).²⁵ STM yielded a topic that fits this definition though there is a strong emphasis on politics. The topic is characterised by “*government*” and “*power*” and involves “*country*”, “*state*”, “*native*”, “*people*”, “*colony*” and “*africa*”. More words in the word cloud for imperialism in Figure (7a) further strengthens our interpretation (“*administration*”, “*interest(s)*”, “*chief*”, “*trade*”, “*military*”). *FREX* lists several words that appear to specifically relate to British imperialism in South Africa like “*annexation*”, “*president-orange*”, “*bechuanaland*”, but there are also broader terms like “*imperial-government*” and “*colonial office*” that could apply elsewhere in Africa. Interestingly, under this topic we find some modest reference to slavery (“*slave*” in *Score* and “*abolition*” in *FREX*).

Travellers frequently interacted with African rulers. In precolonial times, Afro-European relations were often maintained by soldiers, traders and missionaries. Negotiations were often about trade, safe passage and the toll that had to be paid for passing through the territory. Indeed, [Irwin \(1975, p.83\)](#) noted that the “literature of African exploration is full of the exasperated comments of travellers held up by what seemed to them unwarranted delays and exactions.” STM yielded a topic that deals with diplomacy. Words like “*letter*”, “*treaty*”, “*message*” hint at peaceful relations. Instances of sending diplomatic representatives almost always involved centralised states ([Smith, 1973](#)). STM confirms this. We find references to British “*consul-cameron*”, who was appointed to Abyssinia in 1860 and - after a diplomatic faux pas in letter writing - was held captive by Emperor Tewodros II. Similarly, “*king-ashanti*” may be related to the several European dispatches to negotiate safe trade routes to Ashanti.²⁶ There were also numerous instances, where Europeans applied military force to accomplish their political goals. STM indeed extracted a topic describing “*conflict*”. Words include military ranks (“*officer*”, “*major*”, “*colonel*”) and bases (“*camp*”, “*fort*”) and hints at the violent nature of the endeavour (“*attack*”, “*advance*”, “*battle*”, “*fire*”, “*wounded*”).

Conceptually, imperialism turns into colonialism with the presence of significant numbers of permanent settlers from the colonising power ([Horvath, 1972](#)). The African country that was colonised in this sense very early on was South Africa. The two other African countries with sizeable European settler communities by 1900 were Algeria and Rhodesia (today’s Zimbabwe).²⁷ STM returned a topic of

²⁵This definition follows the seminal work of [Horvath \(1972\)](#).

²⁶Between 1816 and 1820 alone there were seven official missions to Kumasi ([Irwin, 1975](#)).

²⁷Around 1900, South Africa had by far the largest share of Europeans (22%). In Rhodesia, European settlers made up

“*European Settlement*”. It covers the means of transport (“*oxen*”, “*horses*”, “*waggon*”, “*road*”, “*railway*”), location (“*cape*”, “*natal*”, “*south africa*”, “*transvaal*”) and economic opportunities (“*gold*”, “*cattle*”, “*farm*”).

The four topics take up considerable space in the documents: Imperialism (6.06%), Diplomacy (6.06%), Conflict (4.76%) and European settlement (2.13%). Topic prevalence over time follows an interesting pattern: Imperialism and Diplomacy see a sudden increase in 1800 and remain important until the end of our study period. In contrast, Conflict was irrelevant in 1600 but then linearly increased over time. European settlement was of little importance until 1750 and then slightly more so later. This pattern is largely consistent with the underlying historical developments. The Portuguese established trading posts in the 15th century and the Dutch East India Company (VOC) created a settlement colony at the cape in 1652. However, for a long time, European powers were unable to control significant parts of African territories (Curtin, 1998), until the second half of the 19th century culminating in the Scramble for Africa 1881. The topical appearance of European settlement after 1800 is probably owed to the British takeover of the Cape Colony in 1795/1806. Finally, Imperialism is a common topic among travellers in South and West Africa, Diplomacy was treated more often in East and West Africa, whereas European settlement is heavily skewed to Southern Africa.

Evangelization, Christian Values & Prayer: During early exploration, clergymen often accompanied voyages. They were also posted to the bigger coastal forts. In that era, their task was mainly to provide ecclesiastical services to Europeans. Evangelization efforts were negligible. The few organized attempts of evangelization were probably discouraged by the very high mortality rates due to malaria (Jedwab et al., 2022). When quinine became a relatively effective malaria prophylaxis in the 1840s, mission societies slowly started to become a significant actor seeking widespread evangelization of African peoples. Missionaries were also expected to report back to their mission societies about their progress and obstacles. Many words in the topic word cloud shown in Figure 7b are related to evangelization efforts during the later phase such as “*missionary*”, “*mission*”, “*work*”, “*station*”, “*society*”, “*sunday*”, “*school*” and “*gospel*”. We also find various Christian denominations such as “*catholic*”, “*wesleyan*” and “*church*” of “*england*”.

The colonial enterprise was often justified with a civilizing mission (Hastings, 1996). Many travellers had a Christian background. They commented on what they observed based on their own Christian value system, which they often perceived as culturally and morally superior. STM returns a topic that we best summarise with the label “*Christian Values*“. We find words that appear to emphasize the family (“*children*”, “*women*”, “*wife*”, “*father*”, “*husband*”, “*marriage*”). The FREX index unambiguously points to a Christian value system, with tokens like “*holy-spirit*” and “*jesus-christ*” and Victorian virtues and vices (“*chastity*”, “*adultery*”) and value judgements (“*righteousness*”, “*sinful*”). The tokens “*learn-read*” and

7% of the population (Easterly, 2003). In Algeria, the French numbered about 5% of the population in 1896 (Stora, 2004). Kenya developed to a settler economy only after 1900 (Jedwab et al., 2017).

“taught-read” probably refers to the bible. It is apparent that in this topic African beliefs and customs (“heathen”, “polygamy”) were likely treated as a juxtaposition to Christian values.

We also find a topic that consists of a collection of biblical English like “thee”, “thou”, “hast” and “lord”, particularly when looking at both *Score and FREX*.²⁸ We can only speculate that travellers in foreign and dangerous lands resorted to prayer. Bible verses can also give a moral justification for judgements and (cruel) actions.

With 6.04% of text, the topic of Christian Values takes a lot of space in the travellers’ writings. The same is true for Prayer (5.17%). In contrast, the topic of Evangelization, which is related but much narrower, amounts to 2.95%. Prayer is highly prevalent in 1700, but then tapers off. The other topics, in contrast, become more important over time, leveling after 1850; both topics were popular among travellers of West and Southern Africa. That Christian Values is more prominent in journeys to Central Africa and Prayers in West Africa appears to us as a hitherto unexplored finding.

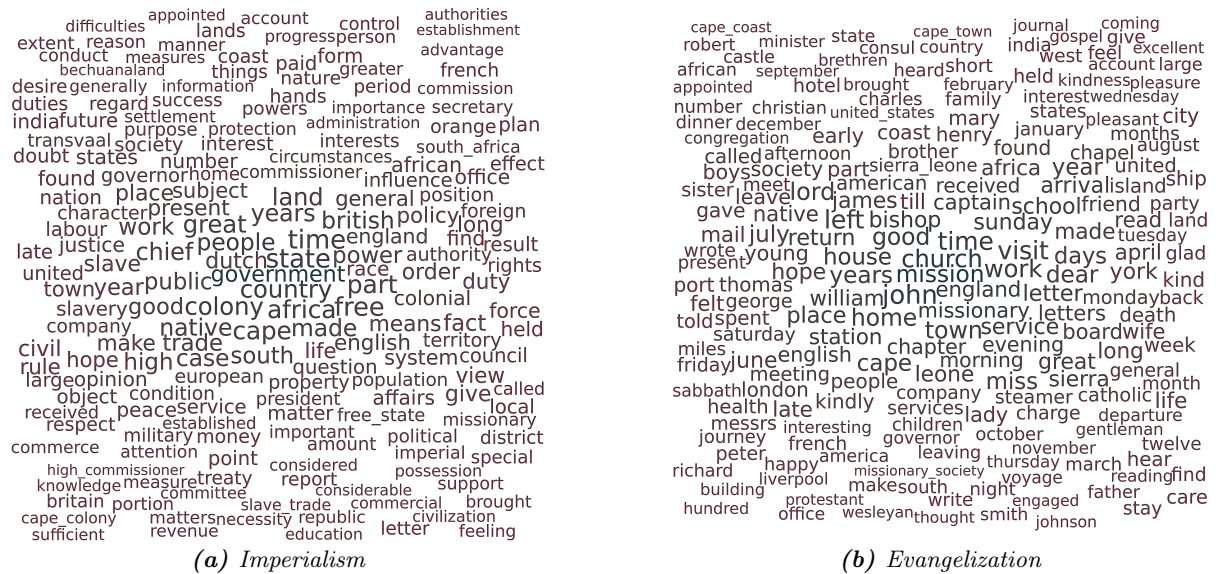


Figure 7: Word clouds encapsulating themes within different topics.

Notes: The word clouds show top 200 words with the highest word prevalence (β), also known as the *Highest Probability* measure. The size of the text is proportional to word prevalence (β). Word clouds of other topics are shown in Online Web Appendix Figures I.

Ethnography: Travellers also frequently made ethnographic observations. “Passages on savage custom, culture and dress” were considered relevant topics in the civilization-savagery debate (Haarhoff, 1989, p. 49). STM identifies these passages remarkably well. Topic words include descriptions of “body” parts like “head”, “hair”, “neck”, “skin”, “legs” as well as details about the “dress” and “wear” such as “feathers”, “cloth” and “beads”. In general, this content is problematic. The 19th century practice of description

²⁸ “Thee, Thou” are common pronouns from the King James Bible translated in 1611 which travellers would be carrying. In principle, these are stop-words. However, the SMART word list did not recognize them as such and did not remove them.

of physical properties and categorisation - seemingly neutral and objective - was a precursor to non-scientific, racial anthropology. The ethnographic topic accounts for 3.94% and was indeed a common element among travellers of any region in Africa.

Climate/Weather, Seafaring & Early Exploration: Travellers were at the mercy of the elements and the dangers that came with seafaring. Hence, the weather topic is dominated by the descriptions of navigation, and the meteorological observations and climatic conditions required for safe navigation. The growth in lexicon evolved by the 16th century when it became easier to determine cardinal directions (i.e. north, east, south, west). More difficult to solve was the longitude problem. Determining how far east or west from a fixed position a ship was required accurately measuring time at sea which was not solved until the 19th century (Weinrit and Neumann, 2017).²⁹ Because of the navigation vulnerability imposed by inclement weather, words like “*overcast*”, “*heavy-shower*”, “*lightning-rain*”, “*storm-rain*”, “*noon-cloudy*” have a higher *FREX*, *Lift* and *Score* index associated with the topic of navigation. Humidity and light both affected the accuracy of the time keeping and hence the ability to locate a position east or west. Furthermore, cloudy skies obscured the stars, and therefore the position of latitude and longitude. When Harrison’s chronometer was more widely adopted in the 1750s we observe an increase in seafaring descriptions.³⁰

The topic of “*Seafaring*” is another pertinent example, with the introduction of the chronometer circa 1750 and the growth in written artefacts about navigation at sea. Notable *Lift* bi-grams such as “*chart-latitude*”, “*longitude-noon*”, and “*weigh-wind*” are examples that enter the lexicon after circa 1750. This is when ships no longer needed to follow the coast for navigation purposes and can use this new technology - the chronometer - to safely navigate. Prior to this, climatic variables and poor time-keeping hindered east and west navigation. The chronometer allowed European explorers and merchants to measure longitude at sea improving trans-oceanic navigation (Miotto and Pascaliv, 2022). We observe a decline from 1800 as improvements in cartography and seabound navigation become a more widely accessible skill for travellers and the then first steam ships were introduced to West Africa in 1852.

The topic of “*Early Exploration*” was also limited to the coastline of Africa, dominated by the Portuguese or Spanish and constitutes 4% of the source texts. This affects the quantum of observations in our corpus as only notable explorers had their accounts translated into English. Texts tend to lack observations about land and people as this cohort was geographically limited to the coast without much contact with the interior. The topic begins around 1450 when Prince Henry of Portugal founded a center for seamanship. Shortly thereafter we start to observe a rise in west African exploration from the Iberian

²⁹A navigator would need two precise timepieces, one set to a standard reference time (the time at the Prime Meridian) and another set to the time of the place where the sailor was. The time of day could be determined at sea by noting when the sun was at its highest point in the sky, i.e., noon. To keep track of the time, a clock was needed; but not even the best pendulum clocks worked at sea. English clock-maker John Harrison perfected his chronometer. Harrison’s chronometer was not affected by motion, temperature, or humidity, and it kept time as accurately on sea as it did on land.“

³⁰[Board of Longitude and the introduction of the chronometer.](#)

peninsula with tokens such as “*king-portugal*” appearing. Early accounts of African exploration around the Cape of Good Hope coincides with the expansion of Portuguese interests to east Asia. From 1511 to 1641 Portugal controlled the strategically important Strait of Malacca, through which all seagoing trade between China and India was concentrated. This is evident in the topic lexicon with the *Lift* terms “*china-malacca*” and “*great-malacca*” produced by the STM. The arrival of the Portuguese also brought about a dramatic shift in Atlantic imperialism and trade (da Silva, 2012). The flow of African gold is evident in the topic words with “*found*”, “*gold*” appearing in the STM’s *Highest Probability* and *Score* index. With the founding of North African outposts, we observe words like “*morocco*”, “*algeriers*” and “*fort*” as Portugal gained access to the trans-Saharan gold trade. Rather than going by caravan to the Italian city-states, the precious metal was shipped by sea directly to Lisbon and western Europe (Klingelhofer, 2010). As Dutch – Portuguese commercial rivalries developed in the Atlantic, we start to see increasing Dutch exploration, and in particular during the 17th century.³¹ According to the STM, the topic predictably phases out around the 18th century when we also note the introduction of accurate cartography and navigation in the prior paragraph.

Health/Medicine & Climate: Once the dangers of the sea voyage were over, the next obstacle confronting travellers was the tropical African climate. Among the many diseases, malaria was the biggest killer (Curtin, 1998; Patterson, 1974). It earned West Africa the name ‘white man’s grave’ (Curtin, 1961; Rönnbäck et al., 2019). High mortality rates restricted travels off the coast. With the introduction of the quinine prophylactic, mortality rates dropped significantly opening the way to the interior.³² The STM topics are dominated by the symptoms (“*delirium*”, “*vomiting*”, “*fever*”, “*disease*”, “*sick*”, “*days*”, “*diarrhoea*”) or diagnosis (“*dysentery*”, “*malarial*”, “*infection*”, “*spleen*”) and then treatment (“*opiate-night*”, “*antimonial-medicine*”, “*drink-opiate*”, “*inoculation*”) of tropical diseases.

Camping: Setting up camp was part of the travelling experience. It took a considerable time. And despite it being a routine task, the conditions of any site may have never been quite the same. The evenings may have left time for reflection (Fabian, 2001). STM identifies a topic, which describes this camping experience. We find words like “*night*”, “*morning*” and “*tent*”. The *FREX index* also points to the simple pleasures of “*evening-meal*” and “*good-supper*” even detailing the cuisine like “*couscous*”, “*fresh-milk*” and “*sour-milk*”. One may infer from words like “*caravan*”, “*camel*” and “*desert*” that the topic would point to travellers in Northern Africa. Surprisingly, however, the topic is more frequent

³¹West Africa had been a monopoly of the Portuguese and then Spanish Crown since the mid-15th century (da Silva, 2012), with the Dutch coming into mercantile conflict from circa 1580 until the Twelve Years’ Truce (1609–21) (Enthoven, 2003).

³²Quinine, the bark of the Peruvian cinchona tree had been used by Europeans since the 16th century. By the 1820s Pierre Joseph Pelletier and Joseph Bienaime Caventou had isolated the active ingredient of the bark (Patterson, 1974). By 1850 soldiers and missionaries used quinine. However, widespread adoption and the proper application of quinine was a gradual process. Jedwab et al. (2022) showed that missionary expansion was related to the adoption of quinine.

with travellers of West, East and Central Africa, which in many parts shares the dry landscape of North Africa. Finally, this topic takes a large share of the writing (5.96%). Temporally, the prevalence largely agrees with the occurrence of land travels, slowly increasing from 1700 and peaking in 1850.

Geology/Geography & Zoology/Hunting: Researching and describing Africa’s geography was a common feature in many expeditions. STM indeed returned a topic that fits ‘geography’ well, with words like “*valley*”, “*mountain*”, “*plain*”, “*rock(s)*”, “*desert*”. FREX returns geological features like “*volcanic*”, “*basaltic*”, “*gneiss*” and geological time scales like “*ecocene*” and “*pliocene*”. The topic has a high prevalence (5.08%); increasing over time and is relatively more concentrated in Northern Africa. In the eyes of European travellers, Africa’s flora and fauna was exotic and so one would also expect the description of the environment to take up a large space. There was also scientific interest in zoology very much like “*Geology/Geography*”. In fact, many travellers returned with specimens for museum researchers. Especially in the 19th century, Africa’s wildlife also attracted big-game hunters from a privileged background like Frederick Selous and Samuel Teleki (Herne, 1999).³³ STM identified a topic that we categorised “*Zoology/Hunting*”.³⁴ In this topic, words like “*shot*” and “*bullet*” appear with a high probability emphasizing the hunting aspect. FREX then lists animals that travellers were writing most about such as “*leopard*”, “*rhino(ceros)*”, “*lioness*”, “*hyena*”, “*antelope*”, “*elephant*” and “*buffalo*”. With 5.43%, “*Zoology/Hunting*” occupies a lot of text. The prevalence increases over time, especially after 1800. Geographically, it is concentrated – with African wildlife – in East, South and Central Africa.

East African Exploration & River Navigation: Interests in East African exploration and river navigation is centred around the European obsession to be the first to find the origin of the Nile, generating a lot of text.³⁵ For example, the STM results have very strong geographic connections to East African places and geography (“*uganda*”, “*zanzibar*”, “*tanganyika*”, “*kilimanjaro*” “*mombasa*”, “*lake*” and “*victoria*”, “*lake albert*”, “*nile*”). The next group of words emphasise expedition descriptions with linkages to famous Nile source explorations (“*expedition*”, “*caravan*”, “*loads*”, “*general grant*”, “*station*”, “*stanley*”, “*masai*”, “*livingstone*”). The importance of Africans in the European exploration of East Africa is also noteworthy. The tropical belt around the Congo and Lake Nyasa was difficult and dangerous to penetrate, therefore required more cooperation with local knowledge than other regions (Shackleton et al., 1972). The topic of “*East Africa Exploration*” with the search for the origins of the Nile is, reassuringly, limited to when it actually took place circa 1848-1876 in our STM epoch estimates (Bridges, 1973).

Trade/Commerce: This was an important topic with 5% of tokens allocated to this context. It also

³³Our database returns 21 travellers accounts with “hunter” as occupation.

³⁴We decided against “Safari”, because it does not convey the fact that 19th century safaris were essentially hunting expeditions.

³⁵The Nile’s source was the last great mystery for European explorers in the 19th century (Bridges, 1973).

suggests that travelling was as much about exploration as the natural endowments linked to trade and commerce. Travellers liked to emphasise the development potential of Africa (Murray, 2016). Haarhoff (1989, p. 47) notes that “explorers were in a sense land surveyors” seeking out land appropriation suitable for trade and commerce. Terminology that describes land endowment such as “cotton”, “citrus”, “soil” and “climate” are prevalent, with production and quantity adjectives like “bushel”, “cultivation”, “large” and “distribution” emphasising this attitude to precolonial exploration. In 1859, the Journal of the Royal Geographic Society noted how travellers were commercially minded and practical in their observations as they sought to explore the interior. This is highlighted in the commercial lexicon which include bi-grams and words like “gold”, “article-trade”, “excellent-quality” and “produce”. Trade was bi-directional, with merchants seeking information about African preferences (Eltis and Jennings, 1988). The growth in trade and commerce is evident in the European seaports, with commercial and geographic societies flourishing, exemplifying the need for written navigation and transaction history. Hamilton (1868) noted how the policing of transatlantic slavery amplified certain types of trade with the coast, and allowed travellers to safely penetrate further inland. The trade lexicon produced by the STM supports this with *Lift* and *FREX* bi-grams like “central-mozambique”, “distribution-upper” and “history-distribution”.

Fertile Crescent, Religious Tourism, Egyptology & Architecture: The Fertile Crescent was a touristic highlight early on. The region became incorporated into the Grand Tour.³⁶ Palestine, in particular, became a popular destination after the restoration of peace in 1840 (Bar and Cohen-Hattab, 2003). Thomas Cook and American Express, both founded in 1860, offered travel packages bringing thousands of travellers to the Middle East. Pilgrimage to the Holy Land flourished. STM returns indeed a topic that can be best described as “religious tourism”. *FREX* lists tokens such as “holy-city”, the Church of the “holy-sepulchre” located at “mount-calvary”, where the “crucifixion” of Jesus allegedly took place. On their journey to Jerusalem and the Near East, travellers stopped by cities like “cairo” and “alexandria”; they followed the “nile” visiting places like “memphis” and “thebes”. Journeys around the world with stops in Egypt became easier with the opening of the “suez” “canal”. Egypt’s ancient cities were also of scientific interest. In addition, the STM yielded a topic that describes the architecture of the cultural and ancient sites: “building”, “temple”, “wall(s)” “tomb”, “pyramids” and “chamber” and were “great”, “small”, “high”, “long”, or of “stone”.

The topics of “Fertile Crescent” and “Religious Tourism” were common in 1600, dropped in 1700 but remained popular 1800-1900. It may be that the two topics reflect more an old form of Christian pilgrimage with a strong spiritual dimension. In contrast, we find the topic of Egyptology and Architecture appearing later, around 1750. The timing coincides with the extension of the Grand Tour and the increased interest in Europe after Napoleon’s invasion of Egypt 1798, which also marked the very creation

³⁶The Grand Tour refers to the educational and leisure trips the of 17th to the 19th century customarily undertaken by privileged, Northern European, young men; originally, Italy and Greece were key destinations.

of Egyptology as a branch of archaeology. The four topics refer to activities that are bound by space. This is confirmed. We find them strongly associated with travels to North Africa. The topics also take a relatively large space in our corpus of documents (5.16%, 3.65%, 2.24% and 3.49% respectively).

Conspicuous Consumption: STM identified a topic where travellers apparently comment on luxuries encountered during their journey. Those are partly related to accommodation as indicated by words like “house”, “room”, “building” and “hotel” often jointly occurring with adjectives like “large”, “beautiful”, “fine”, “gold” and “silver”. In the *FREX* words we also find reference to indigenous rulers such as “prince” and “maharaja”. The topic prevalence is moderate (3.23%), nonexistent in 1700 but expands over time. The geographic focus is Northern Africa, probably encompassing those who travelled further to India.

Publisher & Journal: The two topics of publisher and journal properties are a reminder of what STM does. It picks up frequent co-occurrences of words on a page. Here, STM identified publisher details and journal properties as a topic. This information typically appears on the first ten pages or as header throughout the document. In principle, one can avoid this generic result by adding manually the starting/ending page of the first/last chapter and restricting STM to pages thereafter/before. However, for the time-being, we acknowledge this output, which concerns 0.61% and 1.28% of the text. In the following, we drop these topics from our discussion as they are not of interest to social scientists.

It is also noteworthy what topics did not show up. From the outset of the project, we expected to learn more about slavery. We know that travellers naturally encountered slavery in Africa. They bought and owned slaves, sometimes they commented on slave raiders and captives, e.g. most famously Livingstone.³⁷ We know that travellers came across slave caravans and markets. However, the only topic that contains reference to slavery is Imperialism. We also screened the STM output for words commonly used as synonyms to slave (bond, chattel, thrall, odalisque, escravo/escrava, helot, serf, indentured, servant, attendant, domestic, drudge, handmaid*, lackey, menial), but did not find them listed among any topic. It appears travellers weren’t abolitionists writing long, dense passages about slavery. This warrants more research in the future.

In summing up, STM generally identifies topics in a coherent and extremely meaningful way. For making use of the data source in a temporal perspective, it would be helpful if topics were covered over a large time span. We find that some topics follow notable epochs. These epochs are intuitively descriptive and therefore chronologically rise and decline with underlying relevance. However, many topics exhibit enough temporal and spatial coverage for further quantitative analysis. Some but not all topics appear interesting from an African economic history perspective.

³⁷Kolapo (2000) draws on Clapperton and Lander’s accounts to assess the history of slave raiding and warfare in 19th century Nupe.

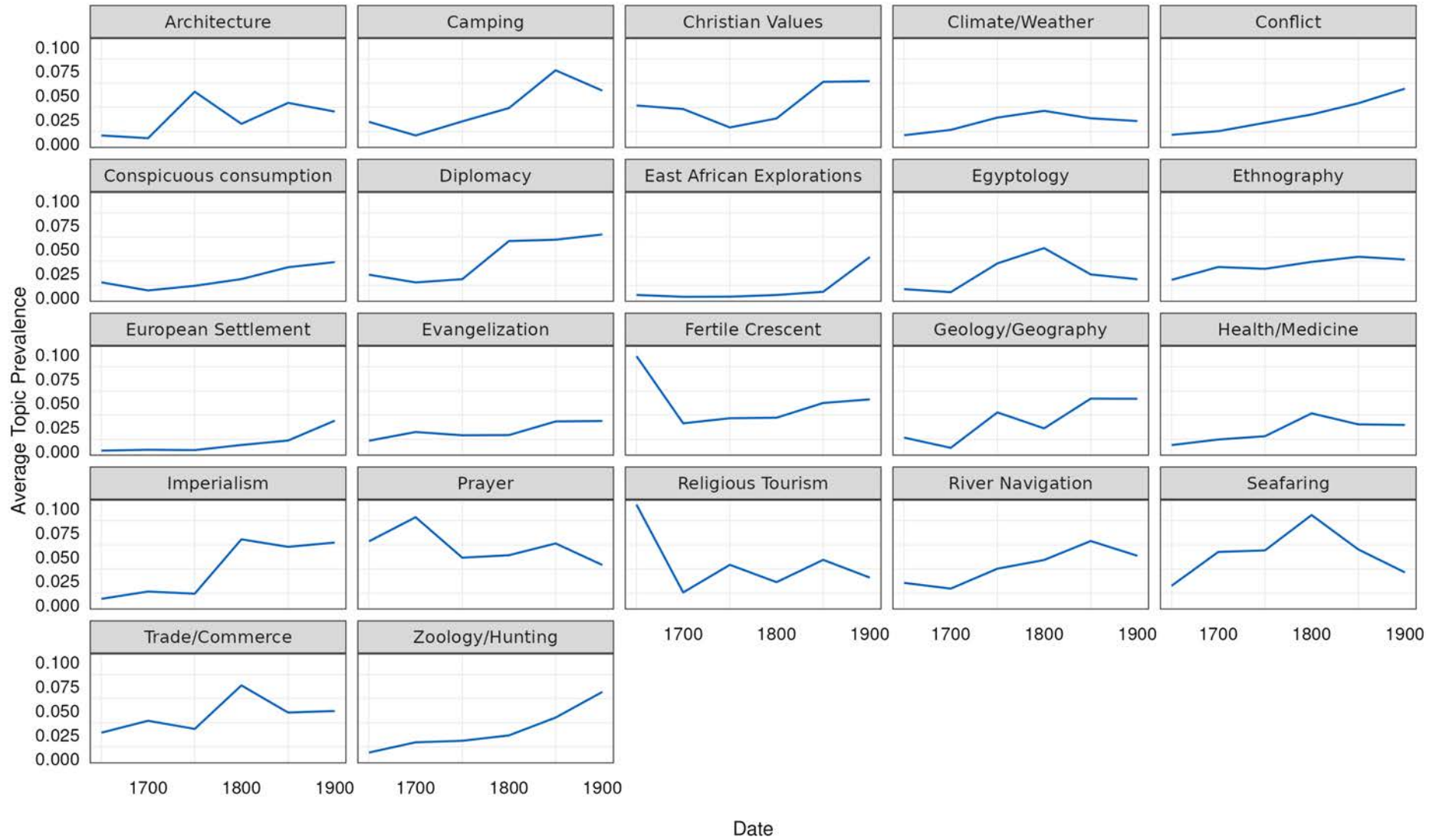


Figure 8: Average topic prevalence ($\bar{\theta}$) over time.

Notes: Parameters are reported in Table 10.

4.2 Topic correlations

One of the key advantages of STM over a traditional LDA specification is that the generative model underlying STM explicitly incorporates the possibility that topic usage is correlated within documents (see equation 6). The topic correlations in our corpus are displayed in Figure 9. Topics are indeed intertwined.

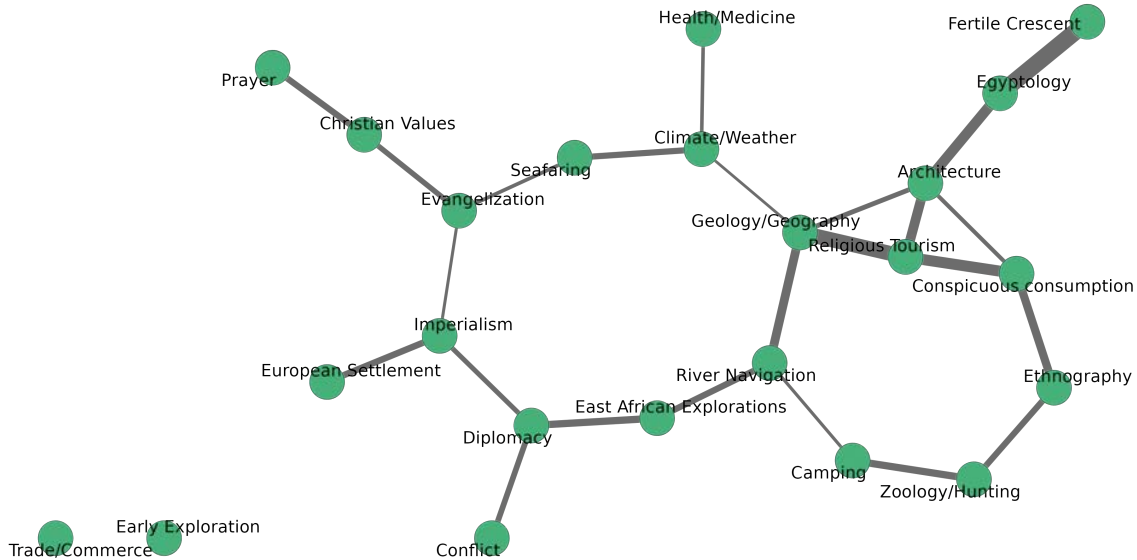


Figure 9: Spearman correlation network between the twenty-five STM topics.

Notes: To keep the network informative we removed topics of limited interest (i.e. Journal Properties and Publisher) with a correlation below 0.01. The thickness of the interconnecting line corresponds to the strength of correlation (min:0.010, max:0.062).

Documents that treat the topic of “*Imperialism*” often include the topics of “*Diplomacy*”, “*European Settlement*” and to a lesser degree “*Evangelization*” as well. Historically, it is apparent why these topics would be correlated. For pursuing European imperial interests, there had to be diplomatic contact with African rulers. Later on, this relationship was formalized with superimposing indirect rule. An approach of direct rule was more common with the presence of European settlers. There was also a symbiotic relationship between the flag and the cross (Stanley, 1990; Andrews, 2009).³⁸ The correlation indicates that in the eyes and documents of the travellers these topics go together (e.g. under an overarching theme of bringing civilization). Similarly, the strong correlation between “*Diplomacy*” and “*Conflict*” points to the proximity of the topics. Diplomacy often precedes and succeeds conflict.

In addition, “*Christian Values*” are correlated with “*Evangelization*” and “*Prayer*”. This correlation is what one would expect to find in writings from the feather of a missionary. Evangelization required additional motivation. It was not the case that saving the soul for the afterlife was considered reason enough. We often find passages that describe local institutions like polygamy and how Christianity would improve the life of Africans. For this, a treatment of “*Christian Values*” is necessary.

³⁸Which one followed is subject to debate and cannot be answered using the STM framework.

It is striking to find “*Ethnography*” correlated with “*Zoology*”. From the early 18th century on, human beings were treated as part of natural history. It became a more and more accepted view that humans were subject to variation like plants or animals, and therefore should be classified in an analogous way (Hudson, 1996). From this, there is only a small step to the “belief that Africans were at least as close to the animal world as to the human world” and a “missing link” between ape and human (Lindfors, 1999, p.8), which would warrant joint treatment.³⁹ “*Ethnography*” is not correlated with “*Geology/Geography*”. Moreover, in our corpus several topic words in “*Ethnography*” hint at racial anthropology. All in all, the uncovered correlation confirms that travellers treated the description of African people document-wise very close to the one of animals.

In addition, we find an association between “*Climate/Weather*” on the one hand and “*Geology/ Geography*”, “*Health/Medicine*” and “*Seafaring*” on the other. Climate and geography both define a country, which is why the two topics are often treated together. Next, travellers indeed made the tropical climate responsible for health conditions (e.g. malaria), therefore they tend to co-occur in a document. Fortune and misfortune in seafaring depends on the weather. It was important information. Hence, the joint treatment in documents.

The very strong correlation between “*Religious Tourism*”, “*Architecture*”, “*Geography/Geology*” and “*Conspicuous Consumption*” strengthens our interpretation of the type of travel, particularly to the Holy Land and Egypt. The fact that these topics accompany each other points to modern day tourism that was not primarily engaged in religious pilgrimage, but sought historical and scenic dimensions and attached great importance on luxuries while traveling (Bar and Cohen-Hattab, 2003). This tourism appears very distinct from travels in the rest of Africa. “*East African Exploration*”, “*Diplomacy*” and “*River Navigation*” are connected through the search for the Nile. European travellers had to traverse through territories around the great lakes controlled by African rulers.⁴⁰ Moreover, the distances, dangers, and frequently changing security situations required local African knowledge (Shackleton et al., 1972). Hence, the increased diplomatic interactions with African rulers. Naturally, the travellers often followed rivers and lakes, hence, the topic of “*River Navigation*”.

Certain topics are thematically isolated, which is interesting in its own right. The topic of “*Early Exploration*”, for example, is set apart by its temporal and spatial dynamics. First, it was dominated by Portuguese travellers during the age of exploration, which impacts the linguistic makeup of the corpus from 1450 until 1700. Second, the topic is influenced by nautical and cartographic observations of the coastal area with limited linkages to land and people of Africa. In addition, “*Trade/Commerce*” appears as a standalone topic. Here, the missing correlation may be owed to the topic’s omnipresence in all the documents.

³⁹At glimpse, we frequently found drawings of animals and Africans in the same document.

⁴⁰The precolonial state of Ganda was preoccupied with controlling the trade routes (e.g. in ivory) and the supply of firearms (Unomah and Webster, 1977).

4.3 Does a traveller’s background affect their accounts?

Can information about the travellers’ background suggest how their travel descriptions shaped world views (i.e. narratives)?⁴¹ In fact, Shiller (2017, p. 46) states that “narratives may be regarded often as causative innovations.” Starting from the incorporation of probabilistic correlation, our STM construct allows us to not only leverage the information contained in the text, but also the information about the text. In this section we showcase the application of metadata integration into the STM within the field of economic history. Specifically the generation of document specific topic prevalence (θ), to question whether occupation affects the breadth of reporting. In addition, we can investigate how topics change over time and space in an econometric framework and whether any statistically significant relationships exist.

One of the narratives often discussed within economic history, is the effect of nationality and primary occupation of colonisers on economic growth and the institutions within a country (Hynes, 1976; Johnson, 2002). It is often said that the British approach to colonisation was to expand trade and commerce, whose travellers were mostly made up of merchants and traders. In contrast, the French approach was to build an empire, where military men and conquest were the main drivers of expansion (Lugard, 1922; Whittlesey, 1937).⁴² To test the hypothesis that background and vocation matter when it comes to such narratives, we estimate whether a traveller’s occupation is a statistically significant factor in modelling the expected proportional amount for a given topic (*Military*, *Missionary* or *Medical*), θ_d , within a document. The topics resulting from the topic modelling exercise may be representative of the occupational composition of the travellers. This begs the question about the breadth of traveller accounts. Did traveller’s with an explicit occupational title write about implicit occupational topics, i.e. did missionaries write about evangelization? Did military men write about conflict, conquest, and politics? Do travel writers conscientiously disregard their vocational training and aim to provide a holistic view of their accounts to their curious readership?

The common approach to estimate the statistical significance of such a hypothesis would be to use an ordinary least square (OLS) regression.⁴³ The problem with this specification is that the variable of interest, θ_d is bounded, $\theta_d \in (0, 1)$, and this violates OLS assumptions. To overcome this, practitioners often transform the variable of interest by applying a *logit* transformation, $\tilde{\theta}_d = \log(\frac{\theta_d}{1-\theta_d})$. The disadvantage of using this transformation is two-fold: (1) the model estimates the mean of the transformation ($\tilde{\theta}_d$) which makes interpretation difficult (Jensen’s inequality) and (2) data is typically heteroskedastic. To address these issues Ferrari and Cribari-Neto (2004) proposed an estimation methodology where the proportion is modeled directly using a suitable parametric distribution, a characteristic which the beta

⁴¹Said (1978) argued that those who collected information (travelled), shaped public perception of imperial lands, and ultimately provided imperialists with arguments to support the extension of European empires.

⁴²See Nettels (1952), and O’Brien (1988) and Gallois (2013) for additional contributions.

⁴³Generalizes to an ANOVA estimation given a single dummy variable as the explanatory variable.

distribution is well suited towards. This model specification is aptly named *beta* regression. The key advantage of this method is that the regression parameters are interpretable in terms of the mean of the original dependent variable θ_d .⁴⁴

The general model is specified as:

$$\begin{aligned}\text{logit}(\mu_{d,k}) &= x_i^T \beta \Rightarrow \mu_{d,k} = \frac{e^{\beta_0 + \sum_{i=1}^p \beta_i}}{1 + e^{\beta_0 + \sum_{i=1}^p \beta_i}} \\ \log(\phi_i) &= z_i^T \eta\end{aligned}\tag{8}$$

where $\theta_{d,k} \sim \mathbf{Beta}(\mu_{d,k}, \phi_k)$. In addition, equation (8) represents the formulation of the variance parameter, known as precision (ϕ). This parameter controls for the dispersion of the data in the model (Ferrari and Cribari-Neto, 2004). Using the general specification, $\mu_{d,k}$ represents the mean proportion of a topic (k) within a document (d) conditional on the occupation of the traveller who wrote the account, the region and the epoch (referenced against all other categories)⁴⁵ as specified by:

$$\theta_{d,k} \sim \mathbf{Beta}(\mu_{d,k}, \phi_k)\tag{9}$$

$$\mu_{d,k} = g(\beta_0 + \beta_k I(\text{Occupation}_d))\tag{10}$$

$$\mu_{d,k} = g(\beta_0 + \beta_k I(\text{Epoch}))\tag{11}$$

$$\mu_{d,k} = g(\beta_0 + \beta_k I(\text{Region}))\tag{12}$$

where $k = 1, \dots, K$, $g(\cdot)$ is the logit link function described in equation 8.

The results of the beta regression returns two parameters: (1) mean and (2) precision denoted by μ and ϕ . In order to obtain the marginal effect, we piece together the coefficient and the intercept:

$$\text{Marginal Effect} = \frac{\exp(\beta_0 + \beta_1)}{(1 + \exp(\beta_0 + \beta_1))} - \frac{\exp(\beta_0)}{(1 + \exp(\beta_0))}\tag{13}$$

$$\text{Baseline Effect} = \frac{\exp(\beta_0)}{(1 + \exp(\beta_0))}\tag{14}$$

$$\text{Proportional Effect} = \frac{\text{Marginal Effect}}{\text{Baseline Effect}}\tag{15}$$

In order to examine the large amount of coefficients estimated from multiple regressions, we visually present the results in Figure 10.⁴⁶ The Figure reports the *proportional effect* of statistically significant coefficients at a 10% level, as per equation (15). If there was no significant difference in the comparison

⁴⁴Estimation of the model was conducted using the *betareg* package developed by Cribari-Neto and Zeileis (2010).

⁴⁵This requires coding the covariate variable as a binary dummy variables.

⁴⁶In total 500 regressions were estimated using 25 topics and 20 unique meta variables.

of the variable of interest with its peers, the proportion would be reported as equal to 1.

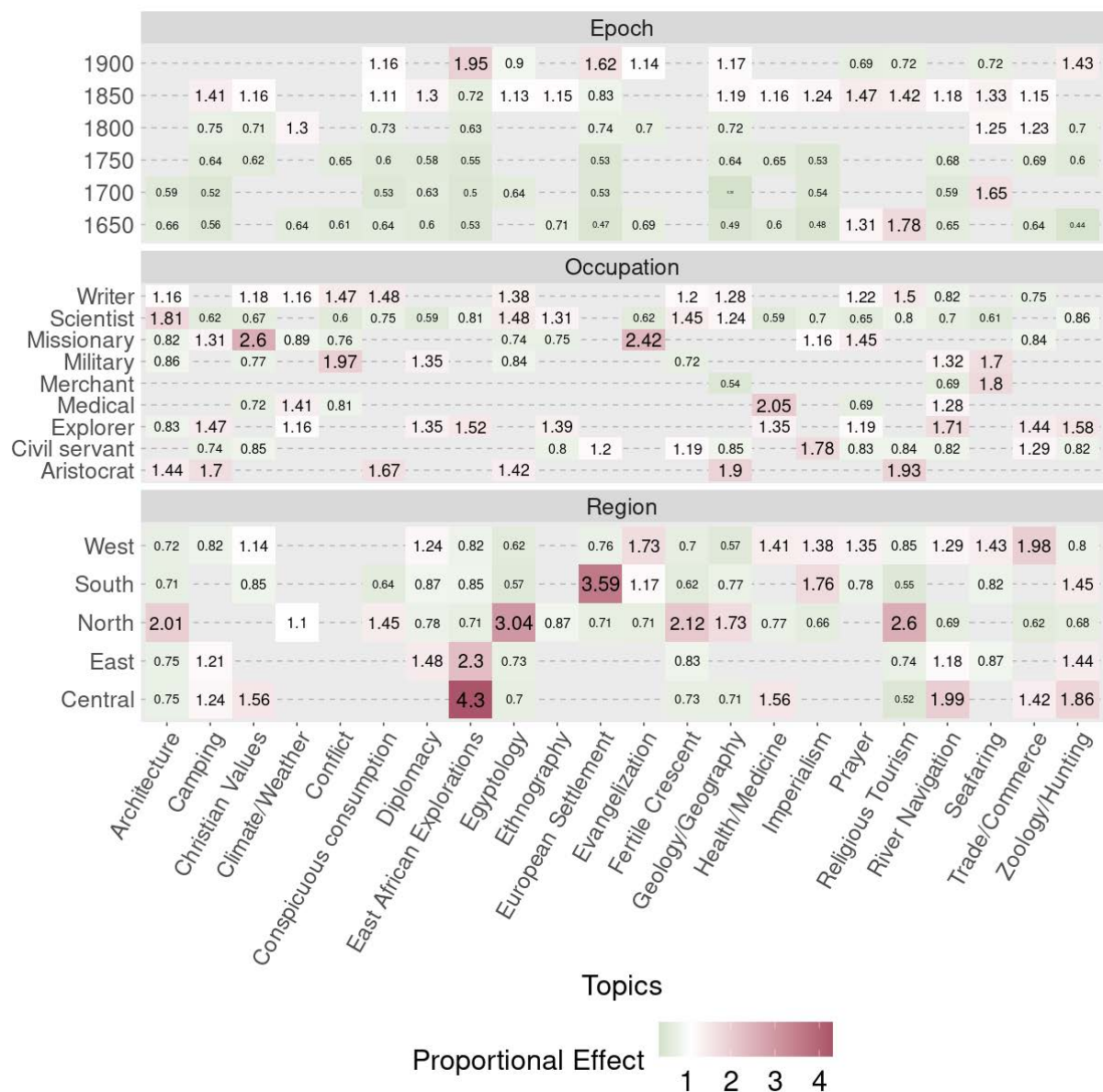


Figure 10: Correlations between topics and time, region and occupation categories.

Notes: This Figure reports the results of the beta regression where proportion of topic, θ_d is modeled as a function of traveller occupation, region and epoch. Results are displayed as the estimated coefficient as a proportion of the baseline estimate. For example the “European Settlement” is 3.59 more prevalent in the southern region of Africa than other regions, while aristocrats discuss camping 1.70 times more often than travellers with other occupational backgrounds. For sake of visual interpretation, we have removed the ‘Early Exploration’ topic as it dominates the 17th and 18th century, with the proportional effect ranging between 4 and 9, which in turn makes visual comparative analysis difficult.

The regression results reveal that in the case where the occupation of the traveller is recorded as being a missionary, their propensity to write about themes that relate to the gospel (as expressed by “Christian Values” and “Evangelization”) within their traveller accounts (expressed as θ_d) is 2.6 and 2.42 times higher when compared to all other occupations respectively. A similar relationship is observed between the travellers with a “Medical” background. They are inclined to discuss topics relating to “Climate/Weather” (possible causes of illness) and cure “Health/Medicine”, while “Military” occupations

have a 1.97 times higher propensity to write about themes around conflict in comparison to their fellow travellers with non-military backgrounds. These findings corroborate the correlation structure found in Figure 9.

The results also reveal that travellers tend to write about a wide array of topics which challenges the hypotheses put forth by Jones (1987) that travel accounts are “one-sided”. The informal contract between travellers, and the consumers of their writings, provide an interesting perspective on how they viewed their obligation to report on fashionable interests back home - offering first hand accounts of what they experienced. By incorporating a large number of narratives from travellers with different backgrounds, we can identify significant changes in attitudes/consciousness towards key topics which deal with the constructive narrative around economic history. Topics that include - but that aren’t limited to - social thought, perceptions, and attitudes, act as crucial predecessors to the formation and thinking of culture, institution and economic development over long spans of time (Said, 1978; Shiller, 2017).

5 Further opportunities in digitizing Africa’s economic history

This paper is part of a larger ongoing research project called **Time Travellers**. The aim is to build a database from traveller accounts: collecting diaries, digitising maps and deploying text analysis. Building on this paper, several steps of work lie ahead.

So far we built a database of written accounts using English search terms (section 2.1). While we showed that the coverage goes beyond English travellers reflecting the engagement of European powers in Africa more broadly (section 2.2), non-English publishing travellers are under-represented. Expanding the search for documents to national libraries using Dutch, German, French and Italian search terms may further broaden the sample of travellers.

In contrast to the traditional topic modeling approach, we added metadata to the topic analysis. The aim was to give greater insight into not only what the text contains, but also the context under which topics were written. The meta variables can be extended and improved. Advancing the spatial meta variables appears particularly promising. We defined travel destinations based on the traveller’s first point of entry using the standardised region meta variable of East, West, North, South and Central. The regions are coarse, spanning large territories. A more precise, geographically finer level would allow for a whole set of new applications that require variation at the local level.

One goal of **Time Travellers** is to link travellers’ descriptions to locations using the traveller’s route maps and place names mentioned in the text. Unfortunately, this will take more time. African place names are extremely difficult to work with. Travellers often lacked African language or phonetics skills. Place names vary in the trade language and local vernaculars creating discrepancies according to who the traveller engaged with. Travellers also resorted to the nearest “home” spelling. As a result, there exists

a large variation in place spellings, in particular across nationalities of travellers.⁴⁷ Two examples noted by Tucker (1948) demonstrate the point. A place in modern day South Sudan was spelled “Regiāf” by Baker (1861), “Rejaf” by Stanley (1873), “Regaf” by Junker (1875) and Casati (1879), and “Reggaf” by Slatin (1879). Similarly, Ouham River in Central Africa was spelled Wam, Wahm, Uam, Ouahm, Ouam, Ouham, and Ouahme. To the best of our knowledge, there is no reference database of historical African place names that would also include settlements that were discontinued.⁴⁸ Even if name variants are known, matching by place name would produce a multitude of possible matches. Travellers often drew their journeys on maps and provided position readings. However, the determination of position based on reference points of unknown coordinates means that we cannot rely on those readings. A solution to this problem is still outstanding.

The final step is making the accounts public as a domain knowledge archive for applications that other researchers and disciplines may want to pursue. Such an archive could provide the basis to flexibly search for historical written accounts in a systematic way. It also reduces the costs in engaging with this source.

6 Conclusion

Using structural topic modelling on a corpus of over 230,000 pages of text we examined 25 key topics in traveller’s documents. STM did astonishingly well identifying topics of interest. Apart from obvious accounts of geography and zoology, we find topics around imperialism, diplomacy, conflict, trade/commerce, health/medicine, evangelization and many more topics of interest to scholarship. We integrated metadata into the model’s generative process to give us greater insight into the spatial, temporal and demographic attributes of the text. We found that the traveller’s background influences the choice of topics. Consequently, the composition of travellers determines topic coverage. We provided first quantitative evidence of the composition. We found a large variation in backgrounds. We also expect this variety to bring different perspective to a topic.

So how useful is a source of traveller accounts? The historiography used the source in two major applications. First, the source is indicative about the travellers and the *Zeitgeist* (epoch) that they were living in. Second, the source can inform us about the territories that individuals visited. Our findings advances the literature in two major respects. First, the literature is selective based on traveller, geography, nationality and topic. We compiled a new exhaustive database of more than 1,971 travellers. Second, the existing literature is qualitative. The findings of this paper showcases how the amalgamation of quantitative methods, state of the art machine learning and domain knowledge could be employed in

⁴⁷ Attempts of standardising English place spellings go back to 1745 (Arousseau, 1941). Over time, several modifications again introduced variation in spellings.

⁴⁸ The gazetteers from NGA (2020) provide name variants that mostly concern changes from colonial to postcolonial times. Using NGA (2020) as a directory of place names produced next to no matches within the diaries.

new applications.

When we started the project, we were concerned that applying an STM would produce a collection of racially abhorrent terms. STM pointed to racist ethnographic passages. At close inspection, derogatory words (like “kafir”) appear in the word clouds, although travellers’ hostile and racist views may be contextual and limited to a specific epoch, nationality and/or region. For example, the words “dark” and “continent” may appear neutral in isolation, but together - indeed a very common metaphor in traveller writings - the expression is problematic (Jarosz, 1992). STM may be unsuitable to pick up the subtle racism in language. On the other hand, STM topic correlations demonstrated how ethnographic observations and zoology were treated in close proximity. In spite of this finding, most topic words appear rather neutral in isolation.

Even though STM cannot provide the semantic analysis needed for deeper contextual analysis, we think that certain words represent factual content and can be interpreted beyond traveller’s subjective value judgements that is attached to it. For example, the occurrence of words like “polygamy” – apparently an obsession of European Christian travellers – contains information and may be possibly used in quantitative applications, particularly if one takes time, space and the traveller’s background into account. Similarly, there may be innovative methods in the future, where the description of “beads” is used to reconstruct market interactions or wealth of communities. Travel diaries are a rich source of information and new advances in text analysis promise to shed new light on Africa’s past 500 precolonial years.

References

- Aitchison, J. (1982), 'The statistical analysis of compositional data', *Journal of the Royal Statistical Society: Series B (Methodological)* **44**(2), 139–160.
- Andrews, E. E. (2009), 'Christian Missions and Colonial Empires Reconsidered: A Black Evangelist in West Africa, 1766-1816', *Journal of Church and State* **51**(4), 663–691.
- Arun, R., Suresh, V., Madhavan, C. V. and Murthy, M. N. (2010), On finding the natural number of topics with latent dirichlet allocation: Some observations, in 'Pacific-Asia conference on knowledge discovery and data mining', Springer, pp. 391–402.
- Arousseau, M. (1941), 'Revision of the R. G. S. II System', *The Geographical Journal* **97**(5), 318–321.
- Austin, G. (2016), Sub-Saharan Africa, in J. Baten, ed., 'A History of the Global Economy', Cambridge University Press, Cambridge, pp. 316–351.
- Bahrami-Rad, D., Becker, A. and Henrich, J. (2021), 'Tabulated nonsense? Testing the validity of the Ethnographic Atlas', *Economics Letters* **204**.
- Bar, D. and Cohen-Hattab, K. (2003), 'A new kind of pilgrimage: The modern tourist pilgrim of nineteenth century and early twentieth century palestine', *Middle Eastern Studies* **39**(2), 131–148.
- Bassett, T. J. and Porter, P. W. (1991), 'From the best authorities': The mountains of Kong in the cartography of West Africa', *The Journal of African History* **32**(3), 367–413.
- Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S. and Matsuo, A. (2018), 'Quanteda: An R package for the quantitative analysis of textual data', *Journal of Open Source Software* **3**(30), 774.
- Berenson, E. (2011), *Heroes of empire: five charismatic men and the conquest of Africa*, University of California Press, Berkeley.
- Blaydes, L., Grimmer, J. and McQueen, A. (2018), 'Mirrors for princes and sultans: Advice on the art of governance in the medieval Christian and Islamic worlds', *The Journal of Politics* **80**(4), 1150–1167.
- Blei, D. and Lafferty, J. (2006), 'Correlated topic models', *Advances in neural information processing systems* **18**, 147.
- Blei, D. M. and Lafferty, J. D. (2007), 'A correlated topic model of science', *The Annals of Applied Statistics* **1**(1), 17–35.
- Blunt, A. (1994), *Travel, gender, and imperialism: Mary Kingsley and West Africa*, Guilford Press.
- Bridges, R. C. (1963), 'The R. G. S. and the African Exploration Fund 1876-80', *The Geographical Journal* **129**(1), 25–35.
- Bridges, R. C. (1973), 'Europeans and East Africans in the age of exploration', *Geographical Journal* pp. 220–232.
- Bridges, R. C. (1982), 'The historical role of British explorers in East Africa', *Terrae Incognitae* **14**(1), 1–21.
- Bridges, R. C. (1987), 'Nineteenth-century East African travel records with an appendix on "Armchair Geographers" and Cartography', *Paideuma* **33**, 179–196.
- Britannica (2022), 'Imperialism', <https://www.britannica.com/topic/imperialism>.
- Chandler, T. (1987), *Four thousand years of urban growth: an historical census*, 2nd edn, Edwin Mellen Press, Lewiston.
- Cribari-Neto, F. and Zeileis, A. (2010), 'Beta regression in R', *Journal of Statistical Software* **34**, 1–24.
- Curtin, P. D. (1961), "'The white man's grave:' image and reality, 1780-1850", *Journal of British Studies* **1**(1), 94–110.
- Curtin, P. D. (1984), *Cross-cultural trade in world history*, Cambridge University Press, Cambridge.

- Curtin, P. D. (1998), *Disease and empire: the health of European troops in the conquest of Africa*, Cambridge University Press, Cambridge.
- da Silva, F. R. (2012), 'Forms of Cooperation between Dutch-Flemish, Sephardim and Portuguese Private Merchants for the Western African Trade within the Formal Dutch and Iberian Atlantic Empires, 1590–1674', *Portuguese Studies* **28**(2), 159–172.
- Driver, F. (1991), 'Henry Morton Stanley and his critics: Geography, exploration and empire', *Past & Present* **133**, 134–166.
- Easterly, W. (2003), A tale of two Kuznets curves: Inequality in the old and new globalizations, in 'Paper for NBER pre-conference on globalization and inequality, Cambridge, MA'.
- Eltis, D. and Jennings, L. C. (1988), 'Trade between Western Africa and the Atlantic World in the Pre-Colonial era', *The American Historical Review* **93**(4), 936–959.
- Enthoven, V. (2003), *Early Dutch Expansion in the Atlantic Region, 1585–1621*, Brill, Leiden, chapter 2, pp. 17–47.
- Fabian, J. (2000), *Out of our minds: reason and madness in the exploration of Central Africa*, University of California Press, Berkeley.
- Fabian, J. (2001), 'Time, narration, and the exploration of Central Africa', *Narrative* **9**(1), 3–20.
- Ferrari, S. and Cribari-Neto, F. (2004), 'Beta regression for modelling rates and proportions', *Journal of Applied Statistics* **31**(7), 799–815.
- Fourie, J. (2016), 'The data revolution in African economic history', *Journal of Interdisciplinary History* **47**, 193–212.
- Gallois, W. (2013), *A history of violence in the early Algerian colony*, Palgrave Macmillan.
- Grajzl, P. and Murrell, P. (2019), 'Toward understanding 17th century English culture: A structural topic model of Francis Bacon's ideas', *Journal of Comparative Economics* **47**(1), 111–135.
- Grimmer, J. and Stewart, B. M. (2013), 'Text as data: The promise and pitfalls of automatic content analysis methods for political texts', *Political Analysis* **21**(3), 267–297.
- Gunn, M. and Codd, L. (1981), *Botanical Exploration Southern Africa*, Flora of Southern Africa, Taylor & Francis.
- Haarhoff, D. (1989), 'Literary ivory: The nineteenth-century travelogue in Namibia and Victorian priorities', *English Academy Review* **6**(1), 42–60.
- Hallett, R. (1963), 'The European approach to the interior of Africa in the eighteenth century', *The Journal of African History* **4**(2), 191–206.
- Hamilton, A. (1868), 'On the Trade with the Coloured Races of Africa', *Journal of the Statistical Society of London* **31**(1), 1–7.
- Hastings, A. (1996), The Victorian missionary, in 'The Church in Africa, 1450-1950', Oxford University Press, pp. 242–305.
- Henderson, M. and Whatley, W. (2014), 'Pacification and gender in colonial Africa: Evidence from the ethnographic atlas', *MPRA Paper No. 61203*.
- Herne, B. (1999), *White hunters: the golden age of African safaris*, H. Holt, New York.
- Hibbert, C. (1982), *Africa explored: Europeans in the dark continent, 1769-1889*, Allen Lane, London.
- Horvath, R. J. (1972), 'A definition of colonialism', *Current Anthropology* **13**(1), 45–57.
- Hudson, N. (1996), 'From "nation" to "race": The origin of racial classification in eighteenth-century thought', *Eighteenth-Century Studies* **29**(3), 247–264.
- Hynes, W. G. (1976), 'British mercantile attitudes towards imperial expansion', *The Historical Journal* **19**(4), 969–979.

- Irwin, G. W. (1975), ‘Precolonial African diplomacy: The example of Asante’, *The International Journal of African Historical Studies* **8**(1), 81–96.
- Jarosz, L. (1992), ‘Constructing the dark continent: Metaphor as geographic representation of Africa’, *Geografiska Annaler: Series B, Human Geography* **74**(2), 105–115.
- Jeal, T. (2007), *Stanley: the impossible life of Africa’s greatest explorer*, Faber and Faber, London.
- Jedwab, R., Kerby, E. and Moradi, A. (2017), ‘History, path dependence and development: Evidence from colonial railroads, settlers and cities in Kenya’, *Economic Journal* **127**(603), 1467–1494.
- Jedwab, R., Meier zu Selhausen, F. and Moradi, A. (2022), ‘The economics of missionary expansion: Evidence from Africa and implications for development’, *Journal of Economic Growth* **27**, 149–192.
- Jerven, M. (2011), ‘A clash of disciplines? economists and historians approaching the African past’, *Economic History of Developing Regions* **26**(2), 111–124.
- Johnson, R. (2002), *British Imperialism, Histories and Controversies*, Palgrave Macmillan.
- Jones, A. (1987), ‘The dark continent: A preliminary study of the geographical coverage in European sources, 1400-1880’, *Paideuma* **33**, 19–26.
- Jones, A. and Voigt, I. (2012), ‘“Just a first sketchy makeshift”: German travellers and their cartographic encounters in Africa, 1850-1914’, *History in Africa* **39**, 9–39.
- Klingelhofer, E. (2010), *First Forts: Essays on the Archaeology of Proto-colonial Fortifications*, Brill, Leiden.
- Kolapo, F. J. (2000), ‘European explorers and aspects of 19th century Nupe history’, *Transactions of the Historical Society of Ghana* **4/5**, 105–122.
- Kopytoff, I. (1988), ‘George Peter Murdock’s contributions to African Studies’, *Behavior Science Research* **22**(1-4), 41–49.
- Kruger, C. E. (2005), ‘Mapping the history of cotton textile production in precolonial West Africa’, *African Economic History* **2**(33), 87–116.
- Lewis, D. D., Yang, Y., Russell-Rose, T. and Li, F. (2004), ‘Rcv1: A new benchmark collection for text categorization research’, *Journal of Machine Learning Research* **5**(Apr), 361–397.
- Linares, O. F. (2002), ‘African rice (*Oryza glaberrima*): history and future potential’, *Proceedings of the National Academy of Sciences* **99**(25), 16360–16365.
- Lindfors, B. (1999), *Africans on Stage: Studies in Ethnological Show Business*, Indiana University Press, Bloomington.
- Lucas, C., Nielsen, R. A., Roberts, M. E., Stewart, B. M., Storer, A. and Tingley, D. (2015), ‘Computer-assisted text analysis for comparative politics’, *Political Analysis* **23**(2), 254–277.
- Lugard, L. F. J. (1922), *The dual mandate in British tropical Africa*, William Blackwood and Sons.
- Matthew, H., Harrison, B., Press, O. U. and Academy, B. (2004), *Oxford Dictionary of National Biography: Brown-Burstow*, Oxford Dictionary of National Biography: From the Earliest Times to the Year 2000, Oxford University Press.
- Michalopoulos, S. and Papaioannou, E. (2020), ‘Historical legacies and African development’, *Journal of Economic Literature* **58**(1), 53–128.
- Miotto, M. and Pascaliv, L. (2022), ‘Solving the longitude puzzle: A story of clocks, ships and cities’, *CEPR Discussion Paper* (No. DP17037), 1 – 47.
- Moradi, A. (2018), Sub-Saharan Africa, in M. Blum and C. Colvin, eds, ‘An Economist’s Guide to Economic History’, Palgrave Macmillan, pp. 285–292.
- Murdock, G. P. (1959), *Africa: its peoples and their culture history*, McGraw-Hill, New York.

- Murray, B. (2016), ‘Building Congo, writing empire: The literary labours of Henry Morton Stanley’, *English Studies in Africa* **59**(1), 6–17.
- Nettels, C. P. (1952), ‘British mercantilism and the economic development of the thirteen colonies’, *The Journal of Economic History* **12**(2), 105–114.
- NGA (2020), ‘Geonet names server’, <https://geonames.nga.mil/gns/html/namefiles.html>.
- Nicholson, S. E., Dezfuli, A. K. and Klotter, D. (2012), ‘A two-century precipitation dataset for the continent of Africa’, *Bulletin of the American Meteorological Society* **93**(8), 1219–1231.
- Nunn, N. and Wantchekon, L. (2011), ‘The slave trade and the origins of mistrust in Africa’, *American Economic Review* **101**(7), 3221–52.
- O’Brien, P. K. (1988), ‘The costs and benefits of british imperialism 1846-1914’, *Past & Present* pp. 163–200.
- Ooms, J. (2021), *pdftools: Text Extraction, Rendering and Converting of PDF Documents*. R package version 3.0.1.
- Ooms, J. (2022), *tesseract: Open Source OCR Engine*. R package version 5.0.0.
- Park, M. and Rennell, J. (1799), *Travels in the Interior Districts of Africa: Performed Under the Direction and Patronage of the African Association, in the Years 1795, 1796, and 1797*, Nineteenth Century Collections Online: Mapping the World: Maps and Travel Literature, W. Bulmer and Company.
- Patterson, K. D. (1974), ‘Disease and medicine in African history: A bibliographical essay’, *History in Africa* **1**, 141–148.
- Pooley, S. (2016), ‘The entangled relations of humans and Nile crocodiles in Africa, c. 1840-1992’, *Environment and History* **22**(3), 421–454.
- R Core Team (2021), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
- Reid, R. (2002), *Political Power in Pre-colonial Buganda: Economy, Society & Warfare in the Nineteenth Century*, Eastern African Studies, James Currey.
- Roberts, M. E., Stewart, B. M. and Tingley, D. (2019), ‘stm: An R package for structural topic models’, *Journal of Statistical Software* **91**(2), 1–40.
- Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., Albertson, B. and Rand, D. G. (2014), ‘Structural topic models for open-ended survey responses’, *American Journal of Political Science* **58**(4), 1064–1082.
- Rockel, S. J. (2006), *Carriers of culture: labor on the road in nineteenth-century East Africa*, Heinemann, Portsmouth, New Hampshire.
- Rönnbäck, K., Öberg, S. and Galli, S. (2019), ‘Working in the ‘white man’s grave’: Wages and migration from Europe to the Gold Coast in the eighteenth century’, *Journal of Migration History* **5**(3), 438–465.
- Ross, A. (2002), *David Livingstone: mission and empire*, Hambledon and London, London.
- Rotberg, R. I. (1970), *Africa and its Explorers: motives, methods, and impact*, Harvard University Press.
- Royal Geographical Society (1844), *General Index to the First Ten Volumes of the Journal of the London Geographical Society*, General Index to the First Ten Volumes of the Journal of the London Geographical Society, J. Murray.
- Royal Geographical Society of London (1839), ‘Proceedings between 1839 and 1880’, *The Journal of the Royal Geographical Society of London* **9**, 1–88.
- Royal Society of Literature (1827), *Transactions of the Royal Society of Literature of the United Kingdom*, Vol. 1, J. Murray.
- Said, E. (1978), *Orientalism*, Peregrine, London.

- Selivanov, D., Bickel, M. and Wang, Q. (2020), *text2vec: Modern Text Mining Framework for R*. R package version 0.6.
- Shackleton, L., Ralling, C., Ingham, K., Simpson, D. H., Bridges, R. C., Marlowe, D. and Attenborough, D. (1972), ‘The search for the Nile sources: Discussion’, *The Geographical Journal* **138**(2), 221–224.
- Sherman, W. H. (2004), ‘Bringing the World to England: The politics of translation in the age of Hakluyt’, *Transactions of the Royal Historical Society* **14**, 199–207.
- Shiller, R. J. (2017), ‘Narrative economics’, *American Economic Review* **107**(4), 967–1004.
- Silge, J. and Robinson, D. (2016), ‘tidytext: Text mining and analysis using tidy data principles in R’, *JOSS* **1**(3).
- Simpson, D. H. (1975), *Dark companions: the African contribution to the European exploration of East Africa*, Elek, London.
- Smith, R. (1973), ‘Peace and palaver: international relations in pre-colonial West Africa’, *The Journal of African History* **14**(4), 599–621.
- Spurr, D. (1993), *The Rhetoric of Empire: Colonial Discourse in Journalism, Travel Writing, and Imperial Administration*, Post-contemporary interventions, Duke University Press, Durham.
- Stanley, B. (1990), *The Bible and the Flag: Protestant missions and British imperialism in the nineteenth and twentieth centuries*, Apollos, Leicester.
- Stiebel, L. (2002), ‘A treasure story: Thomas Baines’s ‘Map to the gold fields of South Eastern Africa’ 1877’, *English Studies in Africa* **45**(1), 1–17.
- Stora, B. (2004), *Algeria, 1830-2000: A short history*, Cornell University Press.
- Supan, A. (1888), ‘Die Fortschritte der Afrikaforschung, 1788-1888’.
- Taddy, M. (2012), On estimation and selection for topic models, in ‘Artificial Intelligence and Statistics’, PMLR, pp. 1184–1193.
- Thornton, R. (1983), ‘Narrative ethnography in Africa, 1850-1920: The creation and capture of an appropriate domain for anthropology’, *Man* **18**(3), 502–520.
- Tucker, A. N. (1948), ‘The Spelling of African Place-Names on Maps’, *Bulletin of the School of Oriental and African Studies, University of London* **12**(3/4), 824–830.
- Unomah, A. C. and Webster, J. B. (1977), *East Africa: the expansion of commerce*, Vol. 5 of *The Cambridge History of Africa*, Cambridge University Press, p. 270–318.
- Wallach, H. M., Murray, I., Salakhutdinov, R. and Mimno, D. (2009), Evaluation methods for topic models, in ‘Proceedings of the 26th annual international conference on machine learning’, pp. 1105–1112.
- Wehrs, D. R. (2008), *Pre-Colonial Africa in Colonial African Narratives: From Ethiopia Unbound to Things Fall Apart, 1911–1958*, Routledge, London.
- Weidmann, C. (1894), *Deutsche Männer in Afrika - Lexicon der hervorragendsten deutschen Afrika-Forscher, Missionare etc.*, Verlagsbuchhandlung Bernhard Nöhring, Lübeck.
- Weinrit, A. and Neumann, T. (2017), *Miscellaneous Problems in Maritime Navigation, Transport and Shipping: Marine Navigation and Safety of Sea Transportation*, A Balkema Book, CRC Press.
- Whatley, W. C. (2021), ‘Up the river: International slave trades and the transformations of slavery in Africa’, *African Economic History Network Working Paper Series* .
- Whittlesey, D. (1937), ‘British and French colonial technique in West Africa’, *Foreign Affairs* **15**(2), 362–373.
- Youngs, T. (1994), *Travellers in Africa: British travelogues, 1850-1900*, Studies in Imperialism, Manchester University Press, Manchester.

A Historical Country and Geographic Name Database

Table 5: Historical names of African regions and territories

Country	Region	Hist Name 1	Hist Name 2	Hist Name 3	Hist Name 4	Hist Name 5
Nigeria	West	Biafra	Hausaland			
Ethiopia	East	Abyssinia	Axum	Eritrea	Ethiopia	Italian East Africa
Egypt	North	Nubia				
DR Congo	Central	Belgian Congo	Congo Free State	Zaire		
Tanzania	East	Tanganyika	German East Africa	British East Africa	Zanzibar	Deutsch-Ostafrika
South Africa	South	Monomotapa	Cape Colony	Batavian Republic		
Kenya	East	British East Africa	Kenya Colony	East Africa Protectorate		
Uganda	East	Equatoria	British East Africa	East Africa Protectorate		
Algeria	North	Barbary Coast	Numidia			
Sudan	North	Nubia	Equatoria			
Morocco	North	Mauritania Tingitana-Morocco	Ifni	Spanish West Africa		
Angola	Central	Reino de Angola	Kingdom of Angola	Ndongo		
Mozambique	East	Monomotapa	Portuguese East Africa			
Ghana	West	Gold Coast				
Madagascar	East	Malagasy Republic				
Cameroon	Central	Kamerun				
Côte d'Ivoire	West	French West Africa				
Niger	West	French West Africa	Hausaland	Sahel		
Burkina Faso	West	Upper Volta	French Upper Volta	Republic of Upper Volta	French West Africa	Sahel
Mali	West	French Sudan	French West Africa	Songhay	Songhai	Sudanese Republic
Malawi	East	Nyasaland				
Zambia	East	Northern Rhodesia				
Senegal	West	French West Africa	Senegambia			
Chad	Central	French Equatorial Africa	Sahel			
Somalia	East	Somaliland	Benadir	British East Africa	Italian East Africa	East Africa Protectorate
Zimbabwe	East	Southern Rhodesia	Monomotapa			
Guinea	West	Guinea				
Rwanda	East	Rwanda-Urundi	German East Africa	Deutsch-Ostafrika	Rwanda and Burundi	
Benin	West	Dahomey	French West Africa	The Slave Coast	Dan Ho Me	
Burundi	East	German East Africa	Deutsch-Ostafrika			
Tunisia	North	Carthage	Numidia			
South Sudan	East	Nubia				
Togo	West	Togoland	Ewe			
Sierra Leone	West					
Libya	North	Western Pentapolis	Numidia			
Congo	Central	Middle Congo	French Congo - Gabo	French Equatorial Africa		
Liberia	West	Grain Coast or Pepper Coast				
Central African Republic	Central	Central African Empire	Ubangi-Shari	Oubangui-Chari		
Mauritania	West	Mauritania Tingitana-Morocco	French West Africa	Sahel		
Eritrea	East	Mdre Bahri				
Namibia	South	German South-West Africa	Monomotapa	Damara Land	Angra Pequena	
Gambia	West	Senegambia	Sahel			
Botswana	South	Bechuanaland	Monomotapa	British Bechuanaland		
Gabon	Central	French Congo - Gabon	French Equatorial Africa			
Lesotho	South	Basutoland	Monomotapa			
Guinea-Bissau	West	Portuguese Guinea				
Equatorial Guinea	Central	Spanish Guinea				
Mauritius	East					
Eswatini	South	Swaziland	Monomotapa			
Djibouti	East	French Somaliland				
Comoros	East					
Cabo Verde	West					
Sao Tome & Principe	Central					
Seychelles	East					
Western Sahara	North	Spanish Sahara				

Table 6: Historical geographic names linked to region keywords and territories

Country	Geo keyword 1	Geo keyword 2	Geo keyword 3	Geo keyword 4	Geo keyword 5	Geo keyword 6	Geo keyword 7	Geo keyword 8
Nigeria	Sokoto	Kano	Lagos	Benue	Yoruba	Niger River	Akassa	Calabar
Ethiopia	Massawa	Harar	Lake Tana	Berbera	Galla Country	Shewa	Danakil	
Egypt	Cairo	Alexandria	Nile	Suez	Memphis	Thebes	Aswan	
DR Congo	Congo	Leopoldville	Katanga	Matadi	Boma	Sangha	Bankana	Stanley Pool
Tanzania	Bagamoyo	Usambara	Kilimandjaro	Zanzibar	Sansibar	Kilwa	Karema	Lake Nyasa
South Africa	Transvaal	Kaffir	Kalahari	Zulu	Natal	Namaqualand	Pondoland	Limpopo
Kenya	Witu	Tana	Kifili	Lake Victoria	Mombasa	Pangani	Baringo	Kikuyu
Uganda	Buganda	Lado	Lake Albert	Busoga				
Algeria	Algier	Sahara	Algiers	Philippeville	Biskra	Bona	Ouargla	Kandouri
Sudan	Darfour	Kordofan	Khartoum	Suakim	Fashoda	Sennaar	El-Obeid	Omdurman
Morocco	Atlas	Fez	Mogador	Tangier	Maroc	Tafilet	Cape Spartel	Marrakesh
Angola	Loango	Benguela	Mossamedes	Malange	Ambrizete	Cabinda	Cassange	
Mozambique	Zambesi	Quilimane	Lorenzo Marquez	Sofala	Chinde	Beira	Kilwa	Maputo
Ghana	Cape Coast	Elmina	Kumasi	Ashanti	Accra	Salaga		
Madagascar	Malagasy	Dauphine	Tamatave	Antananarivo	Fort Dauphin	Amboanio	Morondava	
Cameroon	Doualla	Bipindi	Sangha	Yaoundé	Doume			
Côte d'Ivoire	Grand Bassam	Goree						
Niger								
Burkina Faso	Volta	Ouagadougou						
Mali	Timbuktu	Kayes	Sahara	Segou	Bandiagara	Djenne	Bambouk	Sikasso
Malawi	Lake Nyasa	Livingstonia						
Zambia	Zambezi	Zambezia	Cazembe	Kazembe				
Senegal	Dakar	Saint Louis	Bakel	Carabane	Bambouk			
Chad	Tchad	Lake Chad	Kouno	Chari River	N'Djamena	Logone	Bagirmi	
Somalia	Berbera	Sahel	Zayla	Darror	Mogadishu			
Zimbabwe	Matabele	Rhodesia	Bulawayo	Victoria Falls	Zambezi	Limpopo	Matabeleland	Sabi
Guinea	Conarky	Boke	Gulf of Guinea	Foutah Djallon	Beyla			
Rwanda	Lake Kivu							
Benin	Ouidah	Whydah	Dahome	N'Dali	Bight of Benin	Dahomey	Carnotville	
Burundi	Urundi	Bukeya	Gitega	Bujumbura				
Tunisia	Tunis	Carthage						
South Sudan	Gondokoro	Kotto River	Lado	Bahr-El-Ghazal				
Togo	Lome							
Sierra Leone	Freetown	Falaba	Mayolo	Cape Blanco	Bance Island			
Libya	Tripolis	Tripolitania	Benghazi	Ghadames	Sirte	Ghat	Kufra	
Congo	Brazzaville							
Liberia	Monrovia	Cape Palmas	Musardu	Robertsport				
Central African Republic	Bangui	Obangui	Sangha	Kotto River	Mpoko	Chari River		
Mauritania	Trarza							
Eritrea	Beilul	Asmara	Danakil					
Namibia	Walfishbay	Windhoek	Hereroland	Namaqualand	Swakopmund			
Gambia	Pisania							
Botswana	Okavango	Kalahari	Lake Ngami	Tati River				
Gabon	Libreville	Ashango	Gabon River	Donguila	Batanga			
Lesotho	Basutoland	Morija						
Guinea-Bissau								
Equatorial Guinea	Fernando Po							
Mauritius	Bourbon Island	Port Louis	l'Isle de France					
Eswatini	Pondoland							
Djibouti	Obock	Tadjura	Tadjourah	Danakil				
Comoros	Moroni							
Cabo Verde	Praia							

B Preprocessing of text

For the methods used in this paper, the text of explorer accounts need to be cleaned and transformed into machine readable numbers, also known as *tokenization*.

The preprocessing steps applied are described in detail, accompanied by the number of text that remain after each of the steps

Step 1: Develop software package to OCR pdf files

In the first step, we developed a software package using [R Core Team \(2021\)](#) as our primary statistical tool. Other commonly used text processing packages include [Ooms \(2022\)](#), [Ooms \(2021\)](#), [Benoit et al. \(2018\)](#), [Silge and Robinson \(2016\)](#), [Selivanov et al. \(2020\)](#) and [Roberts et al. \(2019\)](#).

- All pdf's are filename standardised and content-hashed to create a unique identifier per pdf, irrespective of filename.
- Exclude files larger than 100 mb as these contain lots of pictures and very little text.
- Build database structure to house information.
- Extract PDF contents using [Ooms \(2021\)](#) and [Ooms \(2022\)](#).
- Upload text, meta data and other metrics to AWS cloud database. The full database *Entity Relationship Diagram* can be viewed in [Appendix D](#).

Step 2: Tokenization Before tokenization, we have a corpus consisting of 312,781 documents (pages) that successfully were uploaded to the DB.

- covert each OCR word to a machine readable “tokens”

Step 2.1: Pre-processing corpus using traditional data science methods

- Remove punctuation, symbols, numbers and separators etc.
- Only keep English discernible words (66,434 word list).
- Remove stopwords using SMART list.

Step 2.2: Extend Corpus to create bigrams from tokens

- A factor of three for any two word.
- *i.e.* token 1: long, token 2: journey, token 3 bigram: long journey

Step 2.3: Quality Control

- Filter minimum tokens on single page must greater than 100.

Step 4: Covert Tokens to Document Frequency Matrix. In the final step we create the document frequency matrix (DFM). A DFM can be thought of as a summary table of “how many times a word appears in the document”. In our analysis, a document is specified as a single page out of a journal. It is important to note that increasing the number of features has diminishing returns on clustering accuracy. With this in mind we trim the DFM before analysis. Before trimming, the DFM consisted of 237,741 documents; 8,383,715 features features (>99.99% sparse) and 3 docvar, where the document variable describes the region in Africa, the epoch in time and the occupation of the traveller.

- trim DFM to only include the top quantile with minimum term frequency has to be in the top 95
- A large number of tokens has a diminishing return. Don't want infrequent words to be overly noisy

Once the all the preprocessing and trimming steps have been completed the final corpus was reduced to 237,741 documents and 479,647 features.

Step 5: Run the topic model

C Analytical Pipeline

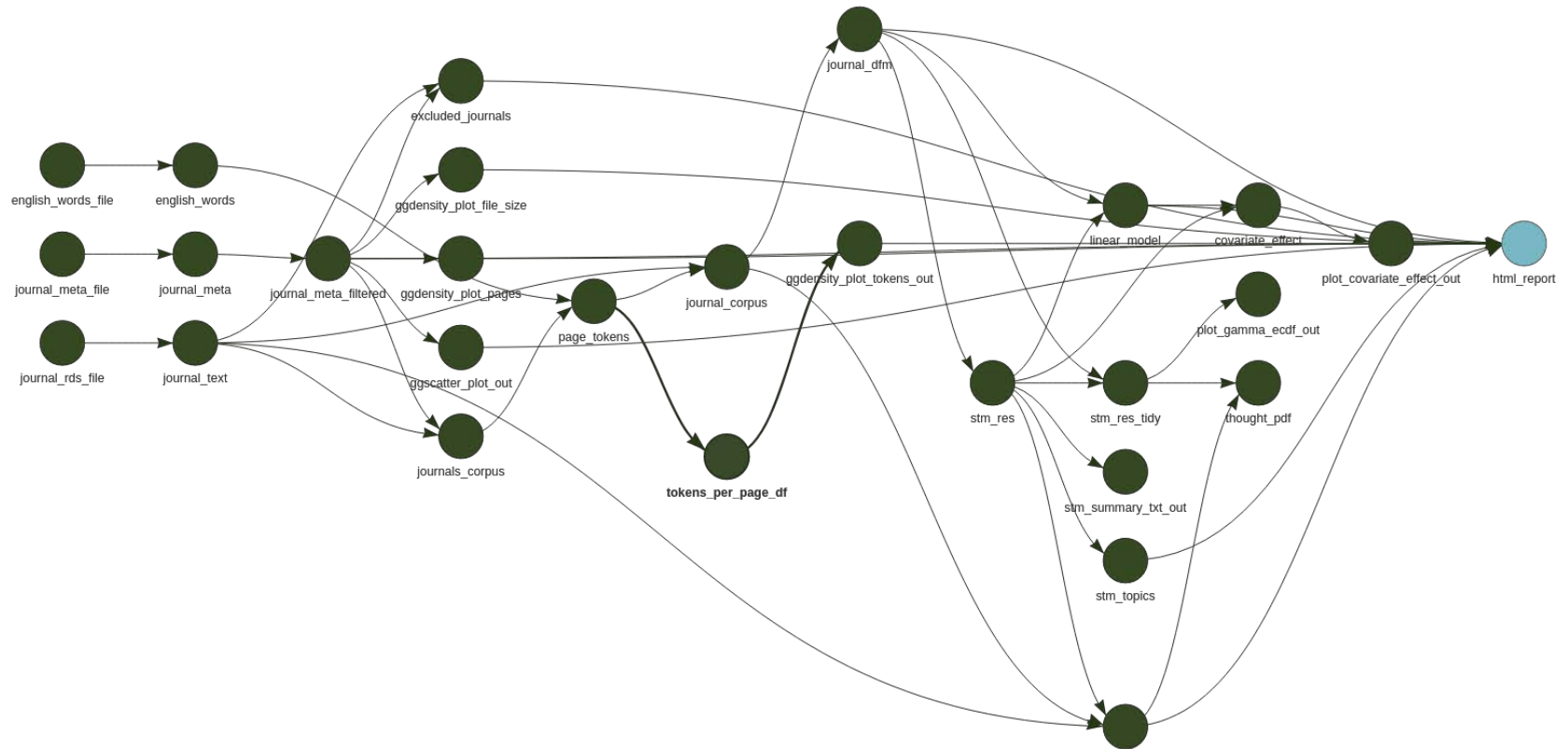


Figure 11: Graph network representation of steps applied before and after the estimation of the STM.

D Entity Relationship Diagram

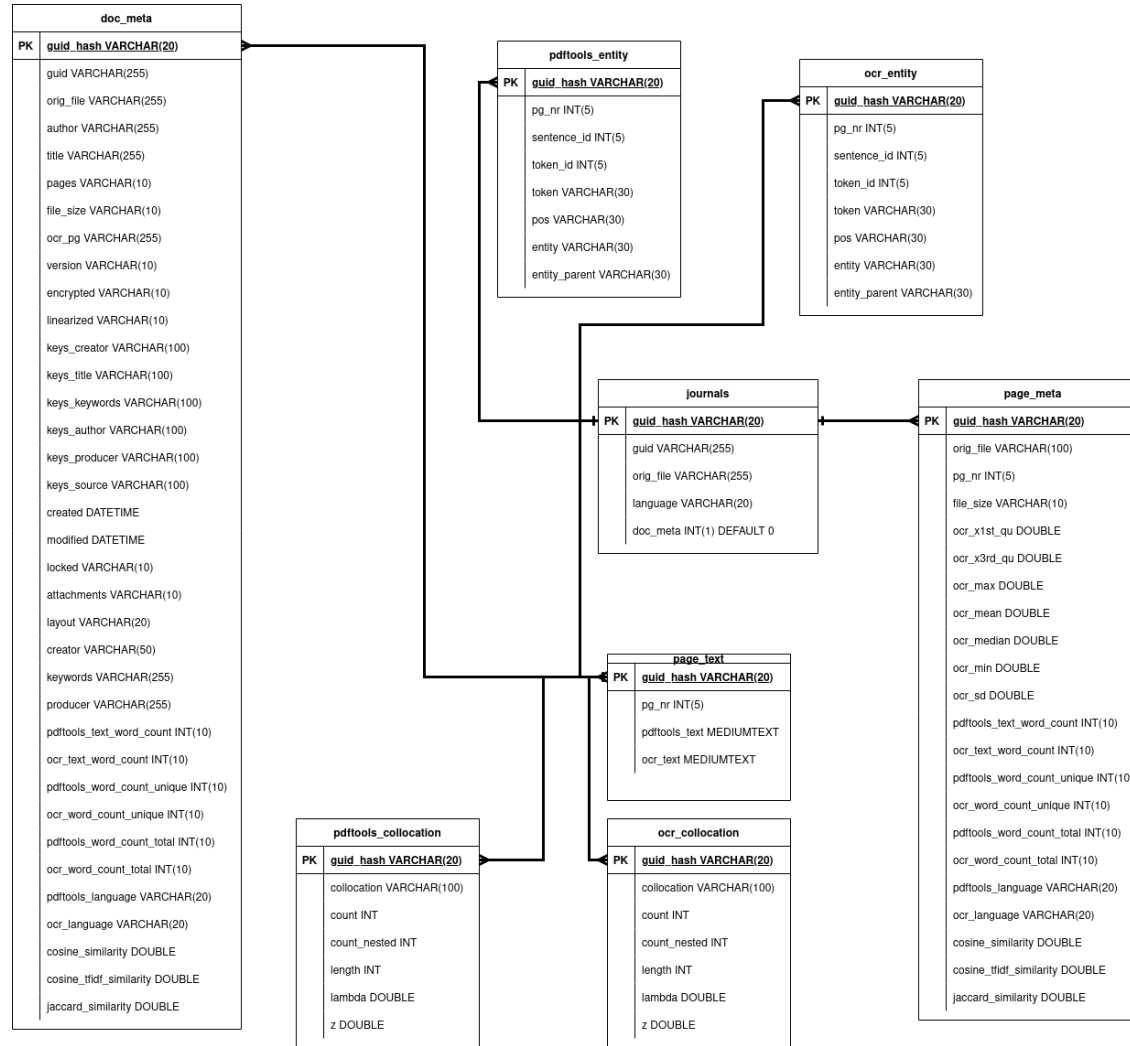


Figure 12: Database structure to support analytical pipeline.

E Topic Selection

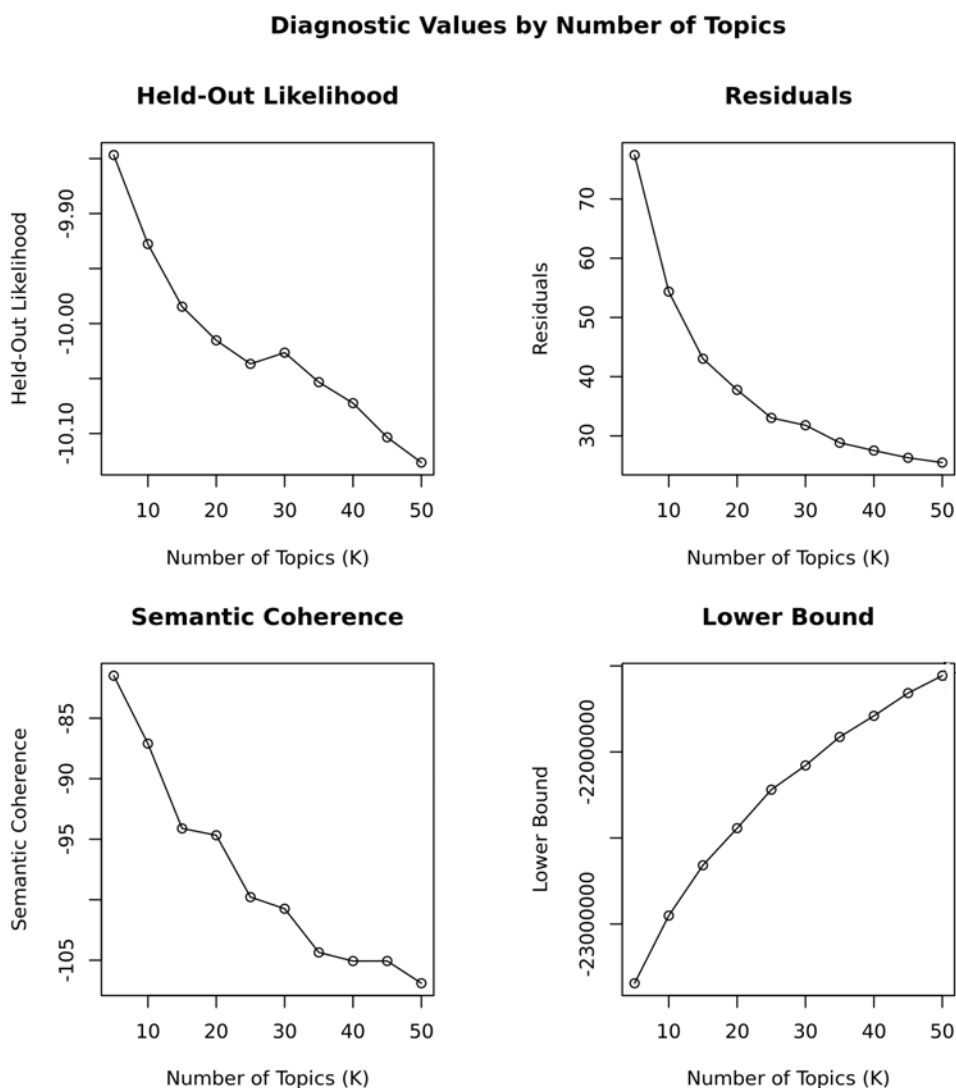


Figure 13: Statistical determination of the number of topics (k) in the estimation process.

F Weighting Schemes in Topic Models

Highest Prob: - are the words within each topic with the highest probability (inferred directly from topic-word distribution parameter β).

FREX: - attempts to find words which are both frequent in and exclusive to a topic of interest. Balancing these two traits is important as frequent words are often by themselves simply functional words necessary to discuss any topic. While completely exclusive words can be so rare as to not be informative. The formula is $FREX = ((w/F) + ((1-w)/E))^{-1}$ where F is the frequency score given by the empirical CDF of the word (w) in it's topic distribution. FREX weights words by their overall frequency and how exclusive they are to the topic

Score: - $\beta_{v,k}(\log \beta_{w,k} - 1/K \sum_{k'} \log \beta_{v,k'})$ where \log probabilities of seeing word v conditional on topic k . Score divides the log frequency of the word in the topic by the log frequency of the word in other topic

Lift: - Lift is calculated by dividing the topic-word distribution by the empirical word count probability distribution. In other words the Lift for word v in topic k can be calculated as: $Lift = \beta_{k,v}/(w_v/\sum_v w_v)$. Lift weights words by dividing by their frequency in other topics, therefore giving higher weight to words that appear less frequently in other topics.

G Topic Words (#30) for the estimated 25-topic STM

1. Climate/Weather (2.47%)

Highest Prob: rain, wind, heat, weather, cold, season, night, heavy, winter, morning, days, storm, time, rains, calm, thermometer, light, winds, months, clear, year, temperature, summer, climate, atmosphere, rainy, thunder, great, evening, lightning

FREX: calm-calm, thunder-lightning, thunder, sere, rain-fell, heavy-rain, slight-surf, wind-rain, epee, pete, rain-thunder, shower-rain, rain-night, fahrenheit, cloudy, lightning, rain, calm-slight, rain-rain, rain-falls, thermometer, surf-calm, drops-rain, overcast, heavy-shower, lightning-rain, storm-rain, eerie, thunder-rain, rain-wind

Lift: calm-medium, calm-slight, cloudy-calm, echt, heat-lightning, medium-calm, noon-dull, overcast-afternoon, pipet, sere-epee, slight-calm, slight-surf, surf-medium, surf-slight, therm-hour, winds-cloudy, calm-calm, medium-slight, paneer, noon-cloudy, reagan, tornado-high, calm-cloudy, surf-calm, lightning-tornado, respiration-evening, tornado-half, lasting-rain, forenoon-clear, amount-ozone

Score: rain, wind, weather, heat, thermometer, calm, thunder, cold, temperature, season, storm, lightning, winds, rains, rainy, calm calm, cloudy, atmosphere, night, winter, heavy, heavy rain, rainy season, slight surf, summer, breeze, thunder lightning, climate, rain fell, noon

2. Journal Properties (1.28%)

Highest Prob: crown, edition, price, cloth, post, author, history, life, works, book, illustrated, english, john, professor, portrait, volume, notes, crown cloth, edition crown, books, extra, series, story, college, numerous, royal, charles, paper, list, gilt

FREX: cloth-price, crown-price, cambric-price, crown-cloth, cloth-extra, illustrated-boards, cloth-limp, post-cloth, edition-cloth, edition-crown, cheap-edition, cloth-boards, alcott, classics-english, limp-cloth, edition-price, square-crown, cloth-gilt, illustrated-cloth, crown-edition, edition-small, extra-gilt, edition-enlarged, illustrated-crown, post-price, library-edition, price-cloth, university-crown, fourth-edition, boards-cloth

Lift: alcott-cloth, american-notes, atlas-history, bacon-part, balfour-philosophical, bleak-house, bound-psalms, bourgeois-type, calf-vellum, cambric-price, carleton-farm, clerical-life, critical-introduction, crown-douglas, crown-gerard, crown-wilson, dictionary-paper, doubles-quits, doyle-leech, dramatic-poem, edition-advanced, edition-cheap, editor-sepoy, eliot-cabinet, gilt-alcott, gilt-cheap, goethe-professor, grant-bart, highland-agricultural, illustrated-boards

Score: edition, crown, price, edition crown, crown cloth, cloth, cloth price, illustrated, cloth extra, post, author, edition post, portrait, crown price, cloth gilt, edition cloth, professor, history, post cloth, gilt, fourth edition, edition enlarged, cheap edition, works, university, college, book, john, large crown, price cloth

3. East African Explorations (2.44%)

Highest Prob: lake, africa, uganda, congo, expedition, zanzibar, stanley, east, country, central, livingstone, coast, station, camp, african, march, people, Nile, ivory, central africa, great, caravan, native, grant, victoria, work, time, masai, chapter, days

FREX: ruwenzori, uganda, lake-albert, albert-edward, tanganyika, kampala, mombasa, lake-victoria, dark-africa, albert-lake, uganda-protectorate, stanley-falls, uganda-egyptian, british-east, king-uganda, mission-uganda, masai, grant-grant, victoria-nile, victoria-lake, swahili, imperial-british, zanzibar, upper-congo, uganda-wilson, bantu, tanganyika-lake, east-africa, nyasa, stanley

Lift: baker-uganda, botany-continued, forest-grant, grant-ruwenzori, plateau-grant, rich-grant, ruwenzori-grant, wilson-baker, wilson-ruwenzori, ruwenzori-uganda, forest-kilimanjaro, grant-uganda, uganda-wilson, wilson-grant, bantu-uganda, ruwenzori-lake, masai-photograph, Nile-masai, ruwenzori-valley, uganda-grant, lake-grant, masai-group, original-bantu, ruwenzori, victoria-grant, speak-bantu, uganda-party, uganda-kingdom, baker-grant, ruwenzori-plateau

Score: uganda, lake, zanzibar, stanley, congo, masai, livingstone, africa, east africa, ruwenzori, tanganyika, central africa, expedition, albert, victoria, grant, bantu, kampala, central, ivory, camp, caravan, lake albert, loads, general grant, Nile, station, kilimanjaro, mombasa, protectorate

4. Conflict (4.76%)

Highest Prob: troops, enemy, army, general, force, camp, time, attack, colonel, left, fire, made, march, wounded, officer, battle, hundred, captain, fort, regiment, command, major, cavalry, position, french, advance, party, front, great, military

FREX: rank-file, mounted-infantry, fire-enemy, native-infantry, heavy-fire, camel-corps, cavalry, file-wounded, infantry, battalion, mounted-troops, attack-enemy, battalion-regiment, regiment, transport-train, detachment, body-enemy, egyptian-cavalry, general-hutchinson, brigadier, commanding-general, khartoum-campaign, left-flank, horse-artillery, advanced-guard, bengal-cavalry, cavalry-artillery, colonel-wood, naval-brigade, cavalry-infantry

Lift: battalion-rank, bivouac-wells, file-rank, file-wounded, square-rear, wounded-rank, baggage-brigade, brigade-line, flank-command, position-battalion, wounded-battalion, brigade-brigadier, file-officer, army-deprived, brigade-advanced, wing-naval, brigade-enemy, major-burn, burn-murdoch, moved-column, russian-works, reserve-command, fourth-brigade, file-missing, drummer-rank, line-brigade, artillery-foot, camp-burns, advance-infantry, foot-battalion

Score: enemy, troops, army, regiment, cavalry, camp, wounded, colonel, infantry, attack, force, brigade, artillery, battalion, battle, officer, major, corps, fire, general, fort, column, garrison, detachment, firing, ammunition, march, captain, shot, fight

5. Health/Medicine (2.73%)

Highest Prob: fever, water, time, disease, sick, days, health, made, great, found, medicine, case, doctor, good, death, patient, medical, pain, food, drink, night, poor, state, suffering, people, lift, left, body, sickness, attack

FREX: lift, quinine, added-lift, infection, dose, attack-fever, calomel, bilious, disease, malady, remittent, delirium, inflammation, fever, intermittent, vomiting, intermittent-fever, dysentery, malarial, severe-attack, spleen, urine, patient, ague, alteration-lift, opiate, sick-list, inoculation, diarrhoea, cough

Lift: demulcent-drink, marine-aged, opiate-night, ordered-bark, repeated-opiate, antimonial-medicine, drink-opiate, flux-continued, gave-copious, lift-marine, ordered-antimonial, ordered-mixture, wine-diet, wine-opiate, lift-varied, antimonial-draught, days-thirtieth, lift-fever, continued-medicine, ounce-bark, list-latitude, lift-complaint, lift-cape, wine-bark, lift-alteration, added-lift, flux-added, fever-lift, complaint-lift, alteration-lift

Score: fever, disease, lift, sick, medicine, medical, patient, health, quinine, complaint, pain, water, added lift, doctor, dysentery, sickness, cure, drink, suffering, illness, poison, tick, attack fever, recovered, attack, malarial, dose, days, stomach, food

6. Zoology/Hunting (5.43%)

Highest Prob: animal, ground, shot, grass, long, wild, found, tree, large, time, feet, water, made, great, round, lion, forest, elephant, small, game, bush, fire, head, back, night, heard, young, make, side, dogs

FREX: leopard, rhinoceros, lioness, carcass, hazel, hyaena, growl, herd, rhino, snake, wild-boar, boar, iris-hazel, stalking, antelope, elephant, buffalo, hyena, carcass, bullet, trigger, wounded-animal, lion, black-rhinoceros, bough, game-central, proboscis, snort, rustling, tiger

Lift: whiff-whiff, calf-lion, charged-bush, warbler-sylvia, wounded-lioness, warbler-warbler, buffalo-alarmed, middle-herd, animal-hard, centre-shoulder, raising-trunk, shot-spine, buffalo-rushed, herd-standing, zebra-meat, thirty-rifle, long-proboscis, herd-roan, found-leopard, life-elephant, directly-shoulder, animal-herd, nervous-force, hazel-river, moment-lion, bird-snipe, blood-spoor, indian-tiger, shot-fourth, count-animal

Score: shot, grass, animal, elephant, tree, lion, bush, rhinoceros, forest, game, feet, wild, herd, rifle, ground, buffalo, bird, antelope, species, leopard, water, fire, dogs, bullet, wounded, legs, beast, jungle, camp, thick

7. Trade/Commerce (4.97%)

Highest Prob: coast, africa, country, species, island, great, found, trade, west, large, fruit, small, tree, soil, part, plant, common, produce, african, parts, interior, land, cape, made, gold, called, portuguese, cotton, leaves, year

FREX: upper-guinea, forestry-west, distribution-upper, forestry, bushels, citrus, lower-guinea, export, wild-common, article-commerce, herbarium, aphis, plant, coast-gold, mozambique-district, guinea-nile, wanderings-west, west-africa, indigenous, superior-quality, article-trade, important-article, soil-climate, flora-tropical, parts-coast, coast-guinea, guinea-mozambique, thrive, mealy, excellent-quality

Lift: africa-oliver, edible-flora, gamble-distribution, guinea-shrub, indian-gamble, island-europe, manual-indian, monograph-distribution, cape-shrub, central-mozambique, commercial-distribution, congo-herbarium, encyclopaedia-industrial, equinoctial-proportion, history-distribution, moore-distribution, oliver-distribution, africa-mann, common-barley, equinoctial-america, plant-recently, common-barren, equinoctial-africa, museum-distribution, uncultivated-gardens, herbarium-congo, botany-distribution, flora-distribution, botany-moore, helena-wild

Score: coast, species, africa, island, plant, trade, fruit, soil, tree, west africa, cape, gold coast, distribution upper, leone, west, sierra leone, sierra, produce, upper guinea, cotton, forestry west, forestry, cultivation, portuguese, guinea, gold, cultivated, country, madagascar, commerce

8. Evangelization (2.95%)

Highest Prob: mission, church, john, missionary, england, time, good, home, sunday, work, bishop, house, service, visit, years, town, station, left, school, william, captain, chapter, james, evening, place, letter, english, arrival, return, cape

FREX: william-cotton, libreville, presbytery, bishop-letters, presbyterian, james-marshall, mission-house, dear-james, sabbath-school, vicariate, church-mission, furlough, honolulu, calypso-july, morning-service, catechist, divine-service, mission-sierra, john-adams, mission-premises, philadelphia, missionary-meeting, saturday-evening, episcopal-church, nassau, american-mission, george-town, wesleyan, methodist, presbyterian-church

Lift: calypso-july, dublin-kelly, james-dublin, barracks-dublin, barracks-lower, brewery-dublin, dublin-james, dublin-john, dublin-joseph, dublin-michael, dublin-miss, dublin-patrick, dublin-thomas, great-dublin, guinness-brewery, hull-england, january-request, john-dublin, john-police, lakewood, miss-richmond, north-dublin, richmond-lunatic, thomas-dublin, william-dublin, dublin-william, miss-dublin, miss-foster, station-dublin, board-goth

Score: church, mission, john, missionary, bishop, william, james, sunday, leone, sierra leone, school, sierra, dublin, miss, chapel, thomas, england, cape, dear, catholic, station, missionary society, society, service, monday, letter, congregation, york, captain, saturday

9. European Settlement (2.13%)

Highest Prob: country, south, miles, cape, natal, south africa, town, oxen, horses, found, road, gold, time, waggon, africa, railway, cattle, good, made, dutch, work, farm, years, great, cape town, colony, wagon, zulu, transvaal, kafir

FREX: pretoria, durban, bulawayo, fort-salisbury, veld, trek, boer, veldt, zulu, pretoria-pretoria, zulu-country, kafir, cart, waggon, port-natal, wagon, natal, team, vehicle, transport-road, land-misfortune, pietermaritzburg, zululand, limpopo, rhodesia, spruit, fort-victoria, trekking, karroo, driver

Lift: perch-pretoria, pretoria-gory, pretoria-found, journey-pretoria, newcastle-pretoria, golden-quarry, days-bulawayo, miles-durban, series-dolomite, gold-commissioner, pretoria-johannesburg, journey-johannesburg, road-durban, robinson-gold, pretoria-great, high-veld, post-boys, johannesburg-pretoria, quartz-reef, country-pretoria, plough-hotel, north-johannesburg, bechuanaland-exploration, durban-miles, journey-durban, johannesburg-town, fields-chapter, trading-stuff, emigration-matabeleland, miles-pretoria

Score: pretoria, natal, south africa, transvaal, waggon, wagon, boer, cape, railway, oxen, zulu, kafir, cape town, kraal, miles, farm, horses, south, durban, dutch, johannesburg, gold, diamond, veldt, port elizabeth, colony, mining, pretoria pretoria, cattle, road

10. Imperialism (6.06%)

Highest Prob: government, country, state, people, colony, africa, time, great, british, native, cape, made, free, years, present, power, public, part, english, chief, general, trade, england, land, south, subject, means, good, work, dutch

FREX: charles-warren, imperial-government, president-orange, austral-africa, austral, executive, colonial-office, colonial-government, captain-bower, legislative, special-commissioner, convention, abolition, bechuanaland, secretary-state, letter-president, high-commissioner, legislative-council, annexation, commissioner-north, legislature, transvaal-government, george-grey, government-cape, government-colony, emancipation, public-opinion, hart-river, governor-cape, commissioner

Lift: agreement-september, annexation-bechuanaland, commissioner-telegram, september-deposition, speedy-annexation, annexation-colony, bower-rhodes, council-treasury, declare-pleasure, instructions-charles, permanent-administration, presence-bechuanaland, prospect-annexation, bechuanaland-high, commissioner-imperial, approval-high, messrs-bower, warren-capetown, warren-hercules, capetown-charles, rhodes-hart, transvaal-flag, bechuanaland-charles, capetown-bechuanaland, imperial-factor, special-commissionership, colony-annexation, cape-opposition, commissioner-hercules, hart-party

Score: government, colony, high commissioner, transvaal, commissioner, free state, bechuanaland, british, cape, state, colonial, orange free, president, policy, cape colony, basutoland, africa, public, south africa, dutch, country, charles warren, trade, free, interests, slavery, imperial government, protectorate, british government, slave trade

11. Camping (5.96%)

Highest Prob: night, morning, people, journey, time, made, village, tent, found, evening, water, hours, left, hour, good, great, brought, party, till, rest, road, gave, camp, long, place, food, back, desert, told, days

FREX: dromedary, sunrise-sunset, tent-pitched, tent, pitch-tent, camel, sour-milk, good-supper, slept-soundly, pitched-tent, door-tent, tent-night, tent-found, evening-meal, tent-door, retired-rest, fresh-milk, ghat, supper, brought-milk, couscous, soundly, hours-rest, bread-milk, night-tent, pass-night, pitching-tent, meal, back-camel, mounted-camel

Lift: sunrise-village, fowl-curry, camel-stolen, minutes-delay, desert-thirst, weary-tree, pretended-asleep, losing-path, sleep-arose, village-wandering, milk-village, saddle-mule, sight-camel, supper-couscous, cold-supper, chicken-dinner, steadily-hours, coffee-tent, time-milking, tent-sheik, chief-meal, sunrise-usual, camel-make, search-oxen, tent-smoke, taking-tent, tent-sand, stood-horse, sunrise-sunset, tent-camel

Score: night, morning, tent, journey, village, caravan, camel, camp, evening, milk, hour, sleep, hours, sunset, people, meal, water, food, party, road, desert, baggage, horses, supper, horse, sunrise, slept, rest, guide, breakfast

12. Early Exploration (4.00%)

Highest Prob: great, king, time, country, made, called, people, fame, part, place, good, make, hundred, land, manner, fort, long, found, water, till, gold, kingdom, account, number, travels, town, large, hath, miles, brought

FREX: thole, reft, fame-time, inhere, whereof, moll, molt, fend, fame, weft, confider, muff, loft, moil, fet-tle, betwixt, fame-manner, boater, caff, travels-discover, king-portugal, mull, foil, part-vincent, thereof, tour-morocco, memoirs-travels, description-algiers, hath, mutt

Lift: great-inhere, muff-pour, thereof-point, ample-relation, bane-great, boater-great, bold-pour, caff-great, caff-miles, caff-point, caff-pour, cape-caff, cape-pour, china-malacca, good-haven, great-banger, great-boater, great-lucre, great-malacca, hall-pour, hath-bane, hath-harbour, hath-point, hold-pour, india-muff, inhere-land, inhere-miles, lading-pepper, land-ball, land-inhere

Score: fame, king, hath, reft, great, weft, thole, fame time, whereof, fort, kingdom, fend, called, gold, thereof, ethiopia, country, moll, inhere, people, molt, loft, cape, cafe, emperor, city, voyage, trade, make, kings

13. Seafaring (4.63%)

Highest Prob: board, boat, island, shore, ship, captain, water, vessel, wind, time, voyage, land, coast, made, morning, night, crew, anchor, port, found, sail, deck, miles, cape, harbour, days, point, long, passage, good

FREX: made-sail, deck, astern, overboard, cast-anchor, dropped-anchor, boat-shore, rigging, captain-weber, starboard, anchorage, pilot-board, anchor, bowsprit, boat-alongside, mainsail, boat-board, shore-boat, thrown-overboard, shoal, obliged-anchor, sail, main-deck, wind-current, deck-found, lying-anchor, boat-boat, offing, amidships, skipper

Lift: isle-bore, search-seal, stood-southward, bower-cable, gale-increasing, wind-starboard, anchor-astern, sail-schooner, clear-wreck, starboard-tack, bring-wind, close-deck, perceived-land, breeze-fell, long-therm, starboard-anchor, fish-vessel, stern-rocks, chart-latitude, port-beam, main-boom, minutes-deck, island-majorca, weigh-wind, deck-heard, rope-board, longitude-noon, light-shore, enormous-shark, anchor-tide

Score: ship, boat, board, shore, vessel, wind, anchor, island, deck, captain, crew, sail, voyage, harbour, water, coast, port, cape, breeze, steamer, ashore, beach, cargo, cabin, canoe, morning, weather, night, brig, anchorage

14. Ethnography (3.94%)

Highest Prob: white, black, head, long, hair, colour, round, brown, women, small, blue, large, yellow, neck, tail, body, wear, dress, length, back, feathers, dark, feet, skin, upper, legs, cloth, made, beads, face

FREX: quill-feathers, quill, yellowish-brown, mandible, tarsi, secondary-quill, anteriorly, upper-tail, primary-quill, blackish-brown, rufous, upper-mandible, dimensions-expanse, length-bill, expanse-wings, brown-feathers, stripe, posteriorly, blackish, rusty-white, brownish, chin-throat, pale-rufous, reddish-orange, round-waist, feathers-white, tail-slightly, reddish-brown, belly-vent, smith-plate

Lift: basin-colour, belly-rusty, belly-surface, bill-compressed, bill-horn, bill-moderately, blackish-purple, brown-base, brown-brownish, cape-adult, centre-feathers, coated-small, dimensions-expanse, edged-reddish, faintly-edged, feather-marked, feathers-base, feathers-blackish, feathers-greenish, feathers-nape, feathers-tawny, figure-moderately, finely-freckled, fourth-quill, frontal-plate, grey-edged, half-collar, intermediate-broccoli, intermediate-reddish, june-colour

Score: air, white, black, tail, brown, feathers, colour, neck, head, yellow, cloth, wear, skin, legs, blue, quill feathers, beads, dress, women, mandible, round, quill, wings, rufous, tarsi, waist, yellowish, long, species, nose

15. Geology/Geography (5.08%)

Highest Prob: valley, mountain, plain, rock, feet, side, rocks, water, road, great, hill, view, miles, desert, mount, south, north, distance, land, part, east, high, place, summit, range, called, lake, level, found, west

FREX: rift-valley, lava, volcanic-rocks, chasm, metamorphic, basaltic, summit-mountain, jordan-valley, volcanic, side-valley, precipitous, side-mountain, eocene, part-valley, crater, metamorphic-rocks, thousand-feet, volcanic-action, scarp, valley-valley, valley-jordan, stratified, pliocene, gneiss, broad-valley, great-valley, deep-valley, narrow-valley, steeply, head-valley

Lift: chasm-ridge, fountain-angle, limestone-series, steam-vent, somali-plateau, cathedral-bastion, fault-scarp, josephus-fountain, northwards-plain, comp-bearings, beautiful-jordan, cretaceous-limestone, angle-terrace, crater-edge, scarp-western, aden-series, border-mountain, volcanic-neck, giant-caldron, zion-west, declivity-ridge, alabaster-basalt, upper-desert, formation-rift, columnar-lava, eastern-view, deccan-traps, north-lebanon, ridge-castle, valley-southwest

Score: valley, mountain, plain, rock, rocks, summit, feet, steep, lake, ridge, hill, mount, road, rift valley, volcanic, sinai, sandstone, rocky, range, level, jordan, water, desert, side, miles, ascent, limestone, ravine, plateau, ruins

16. Conspicuous Consumption (3.23%)

Highest Prob: large, house, room, palace, small, gold, great, silver, made, covered, building, walls, open, round, rooms, english, built, fine, side, floor, beautiful, square, prince, door, court, garden, place, hotel, roof, table

FREX: wales-tour, tour-chap, rooms, prince-wales, verandah, ground-floor, thatched, maharaja, jewellery, large-room, inlaid, furniture, european-style, sherbet, windows, thatched-roof, house-large, electric-light, latticed, mikado, tray, raja, arabesque, floor-covered, bedstead, glass-windows, grand-hotel, stalls, patio, restaurant

Lift: opening-courtyard, drawing-paul, scarab-scarab, palace-lined, gold-lacquer, corner-rooms, canopy-covered, maharaja-prince, gilt-arabesque, cuisine-choice, lacquer-work, advertiser-hotel, hotel-electric, verandah-side, handsome-gilt, good-cuisine, hotel-finest, gilt-frames, windows-coloured, proprietor-manager, scarlet-umbrella, modern-comfort, pretensions-architectural, ejector, garden-handsome, walls-cool, japanese-ladies, beautifully-paved, rooms-twenty, drawing-gates

Score: house, palace, room, rooms, hotel, floor, large, windows, silver, gold, walls, roof, marble, covered, building, silk, wooden, square, door, court, street, prince, built, garden, cloth, gilt, apartment, furniture, glass, dinner

17. Egyptology (2.24%)

Highest Prob: egypt, cairo, Nile, alexandria, egyptian, canal, china, chinese, time, great, nubia, suz, city, thebes, upper, made, desert, embassy, country, found, ancient, part, upper egypt, year, arab, place, delta, present, viceroy, pasha

FREX: embassy-china, upper-egypt, rosetta, chinese, barrage, alexandria-cairo, canal, cairo-alexandria, voyage-nile, lower-egypt, embassy, visit-egypt, cairo, Nile-cairo, egypt-nubia, travels-upper, nubia, canton, suz-canal, cairo-suez, drachmae, nome, banks-nile, return-cairo, miles-cairo, egypt-cairo, canal-nile, alexandria, egypt-nile, noun

Lift: department-scene, scene-superintendent, father-scribe, nome-biography, astronomical-london, direction-lords, auxiliary-method, clearing-lunar, royal-lunar, transit-circle, astronomical-results, stars-description, lucius-greeting, plate-folio, lunar-royal, greenwich-compared, chinese-interpreter, egypt-lotus, drachmae-total, years-scar, centre-astronomical, embassy-chinese, folio-volume, folio-nautical, folio-sexagesimal, jupiter-compared, nautical-table, reduction-astronomical, sexagesimal-table, stars-folio

Score: egypt, cairo, Nile, canal, alexandria, chinese, thebes, china, egyptian, embassy china, embassy, nubia, upper egypt, suz, lower egypt, canton, viceroy, rosetta, memphis, delta, nome, city, egypt nubia, ancient, pasha, upper, temple, suz canal, inundation, antiquities

18. Religious Tourism (3.65%)

Highest Prob: city, town, place, church, holy, built, jerusalem, walls, mosque, gate, called, great, ancient, ruins, convent, house, large, time, greek, stone, wall, small, visit, hundred, mount, christian, sultan, spot, building, side

FREX: item, holy-sepulchre, calvary, holy-city, mosque, church-holy, monastery, bethlehem, sepulchre, minaret, crucifixion, item-town, jaffa, convent, visit-holy, latin-convent, great-mosque, city-gate, jerusalem, mount-calvary, hospice, mosque-built, high-altar, city-walls, greek-church, walls-city, empress-helena, annunciation, golden-horn, mount-sion

Lift: guilt-item, indulgence-years, item-possess, years-item, item-chapel, place-absolution, item-temple, part-forgiveness, absolution-item, item-league, hundred-indulgence, forgiveness-item, absolution-forgiveness, franchise-item, item-castle, item-villa, item-monastery, item-continued, sing-village, altar-person, item-head, item-close, palace-item, league-item, mountain-item, item-mountain, item-village, town-item, spot-virgin, item-franchise

Score: city, jerusalem, mosque, town, church, holy, convent, item, walls, gate, built, sepulchre, ruins, holy sepulchre, chapel, greek, holy land, damascus, tomb, monastery, ancient, bethlehem, altar, wall, place, marble, temple, jaffa, tower, gates

19. River Navigation (5.51%)

Highest Prob: river, miles, water, country, bank, village, stream, large, called, banks, place, small, distance, great, town, side, left, north, feet, journey, people, found, south, road, days, half, mile, high, march, deep

FREX: river-river, bank-river, banks-river, river-miles, left-bank, miles-river, north-bank, small-river, river-wide, main-stream, river-banks, south-bank, side-river, river-found, part-river, river-running, branch-river, river-large, river-hundred, river-point, river-opposite, close-river, river-time, distance-river, cross-river, main-river, river-stream, large-river, river-broad, river-water

Lift: mayo-place, bank-mayo, river-faro, informant-forgotten, lodging-farm, beginning-rocky, village-mountainous, girt-fine, group-shallow, confluence-quango, affluent-affluent, island-kora, inhabited-conjointly, flowing-rate, unknown-march, short-itinerary, feathery-bristle, volume-cubic, middle-flowing, rain-front, current-dashed, march-farm, ninth-hours, considerable-creek, keeping-sandy, fissure-made, tenth-hours, grassy-island, gray-slate, residence-rebel

Score: river, miles, water, bank, stream, banks, village, country, distance, mile, feet, lake, bank river, river river, left bank, niger, north, town, Nile, side river, canoe, road, rapids, called, banks river, district, journey, grass, side, sandy

20. Fertile Crescent (5.16%)

Highest Prob: egypt, history, great, years, time, work, ancient, egyptian, language, present, century, world, country, period, made, french, empire, book, account, year, written, part, greek, life, character, general, subject, found, knowledge, king

FREX: history-egypt, assyria, semitic, egyptology, babylonian, ottoman-empire, revue, mesopotamia, line-read, prussia, manuscript, roman-africa, roman-empire, present-work, chronology, phoenicia, shepherd-kings, monarchy, austria, archaeology, orthography, historian, isthmus-suez, euphrates, progress-egyptology, lists, hebrews, assyrian, punic, present-volume

Lift: automated, google-book, knapsack-guide, post-knapsack, copyright-holder, domain-book, domain-books, domain-materials, domain-united, environs-post, lines-certificate, vesuvius-post, coptic-literature, permission-copyright, provence-pyrenees, pius-complete, leon-asturias, madrid-basque, project-license, terms-project, berlin-publication, project-works, vocalization, uncial-hand, ptolemy-born, imperfect-lines, coptic-studies, hebrew-accounts, reverend-birth, exodus-building

Score: egypt, history, dynasty, century, ancient, egyptian, empire, greek, reign, rome, hebrew, modern, language, roman, historical, book, kings, papyri, years, literature, france, written, period, religion, spain, syria, text, french, asia, conquest

21. Architecture (3.49%)

Highest Prob: feet, temple, side, wall, stone, found, great, walls, tomb, part, small, high, long, large, remains, building, figures, height, chamber, pyramid, length, entrance, built, form, made, granite, rock, square, time, egyptian

FREX: pylon, architrave, cornice, entablature, chamber-feet, main-chamber, lines-hieroglyphics, peristyle, covered-hieroglyphics, portico, chamber-tomb, pedestal, winged-globe, front-temple, side-chambers, crude-brick, great-temple, false-door, casing, pyramid, high-thebes, granite-sarcophagus, abacus, small-temple, pyramid-built, cartouche, lateral-chambers, monolith, temple-small, south-wall

Lift: figure-porcelain, finely-high, hall-measures, memphis-porcelain, obelisk-plan, osiris-underneath, passage-horizontal, porcelain-memphis, portico-figures, statue-calcareous, stone-female, tablet-osiris, unfinished-small, walls-ylon, doorway-ylon, height-umbilicus, lateral-wall, tomb-shaft, hall-closed, pavement-ylon, goddess-side, figure-calcareous, underneath-figures, chambers-dimensions, colossi-size, pylon-feet, portico-ylon, statue-plan, form-facade, grand-passage

Score: temple, feet, pyramid, wall, tomb, hieroglyphics, thebes, stone, walls, chamber, granite, inscription, statue, osiris, chambers, side, dynasty, figures, portico, rameses, pylon, feet high, ruins, height, rock, building, sculptured, horus, feet long, sarcophagus

22. Diplomacy (6.06%)

Highest Prob: letter, chief, king, people, country, time, received, made, governor, return, government, british, arrival, send, present, great, place, english, order, told, make, give, expedition, abyssinia, peace, french, letters, orders, consul, pasha

FREX: british-resident, letter-british, consul-cameron, major-warden, mahdi, resident-secretary, warden, resident-chief, debra-tabor, captain-cameron, letter-king, king-theodore, received-letter, king-abyssinia, answer-letter, letter-assistant, british-abyssinia, resident-high, emperor-theodore, letter-governor, letter-majesty, commissioner-bloemfontein, british-consul, king-ashanti, chief-british, warden-letter, secretary-high, letter-queen, receipt-letter, royal-camp

Lift: resident-june, majesty-honourable, russell-letter, boyd-governor, majesty-abyssinia, dyke-british, friendship-abyssinia, relative-mahdi, resident-february, resident-december, letter-theodore, samuel-steward, consul-despatch, blanc-cameron, cameron-party, cross-seal, abyssinia-august, cameron-emperor, release-consul, letter-matter, residency-february, interview-theodore, letter-comte, lesseps-letter, resident-july, rebellion-chap, restitution-made, warden-extract, theodore-reaches, letter-surrender

Score: letter, king, mahdi, chief, british resident, governor, government, british, consul, pasha, received, treaty, send, letter british, gordon, people, abyssinia, sultan, arrival, theodore, return, sheikh, mohammed, excellency, message, peace, majesty, news, country, orders

23. Prayer (5.17%)

Highest Prob: king, time, told, good, great, made, people, house, hand, lord, gave, give, thou, friend, thought, heard, make, hands, death, long, father, poor, life, brother, present, young, found, thee, word, words

FREX: pepys, thou, thou-hast, hast, thee, shalt, thou-shalt, thee-thou, thou-thou, thine, didst, hast-thou, thou-wilt, thyself, dost, shook-hands, creek-town, wilt, lady-hester, thou-didst, hewer, hester, clinton, thee-thee, give-thee, rupert, braw, shalt-thou, duke-town, lord-thou

Lift: pleased-beseech, pedro-fear, waters-thee, pepys-duke, cantor-account, weep-cantor, servant-pepys, king-yard, lord-rochester, pepys-march, thee-listen, kill-thee, pepys-hewer, listen-thee, pepys-pepys, hewer-pepys, pepys-september, hill-pepys, miss-lyle, picnic-cape, subject-pepys, water-frog, pepys-honoured, maria-thou, ambassador-word, frog-water, pepys-colonel, page-pepys, pepys-vincent, message-ambassador

Score: thou, king, thee, pepys, lord, told, house, friend, father, gave, brother, thou hast, heard, hand, prince, conversation, spoke, death, lady, thought, knew, give, heart, hast, wife, words, good, poor, voice, hear

24. Christian Values (6.04%)

Highest Prob: people, children, life, women, great, good, woman, wife, time, christian, young, things, world, death, love, work, made, make, religion, father, years, spirit, child, read, faith, christ, poor, heart, wives, family

FREX: polygamy, divorce, bridegroom, husband-wife, holy-spirit, jesus-christ, righteousness, sinful, bride, marry, taught-read, divorced, marriage, evil-spirits, holy-ghost, christ-jesus, sinner, salvation, believer, child-born, chastity, love-christ, future-state, heathen, adultery, love-love, children-children, rabbi-mendel, learn-read, mendel

Lift: positive-holiness, humanity-christ, means-belief, pagan-lands, precept-precept, heart-changed, girl-husband, bridegroom-father, saviour-chapter, matter-woman, evening-bunyan, faith-baptism, leave-husband, advanced-thinker, faith-heard, human-held, married-divorced, heart-live, attend-sundays, plenty-dash, christianity-resolved, father-girls, give-expensive, worship-nature, gospel-taught, marriage-purchase, koran-wives, bind-bind, days-consummation, consent-marriage

Score: christ, jesus, children, gospel, people, religion, women, christian, woman, love, faith, wife, child, life, church, holy, wives, worship, jesus christ, father, husband, spirit, marriage, religious, mother, soul, prayer, heathen, read, christianity

25. Publisher (0.61%)

Highest Prob: london, society, africa, south, journal, town, cape, african, geographical, part, royal, great, country, made, magazine, time, south africa, notes, missionary, found, geological, years, report, good, paris, geographical society, visit, journey, cape town, transactions

FREX: society-london, geographical-society, geological-society, royal-geographical, magazine, journal-geological, decade, zoological-society, geological, quarterly-journal, london-missionary, transactions, london-london, royal-society, london, professor-owen, british-association, london-paris, medal, geological-magazine, monthly-magazine, part-london, crania, society, geographical, quart, zoological, cape-monthly, falconer, journal-royal

Lift: colliery-guardian, decade-london, magazine-decade, science-report, dublin-philosophical, geologist-london, london-geological, province-monthly, transactions-geological, journal-geological, africa-quarterly, african-philosophical, transactions-south, london-proceedings, cape-monthly, arts-london, geological-society, geological-magazine, london-notes, africa-transactions, province-magazine, series-cape, association-advancement, philosophical-magazine, philosophical-society, london-coal, african-geology, journal-london, society-london, guardian-london

Score: london, society, africa, geographical, geographical society, geological, society london, journal, cape, south, south africa, royal geographical, magazine, african, geological society, cape town, geology, london missionary, missionary society, missionary, transactions, journal geological, town, royal, association, quarterly journal, south african, paris, london london, notes

H Web Appendix - not for publication

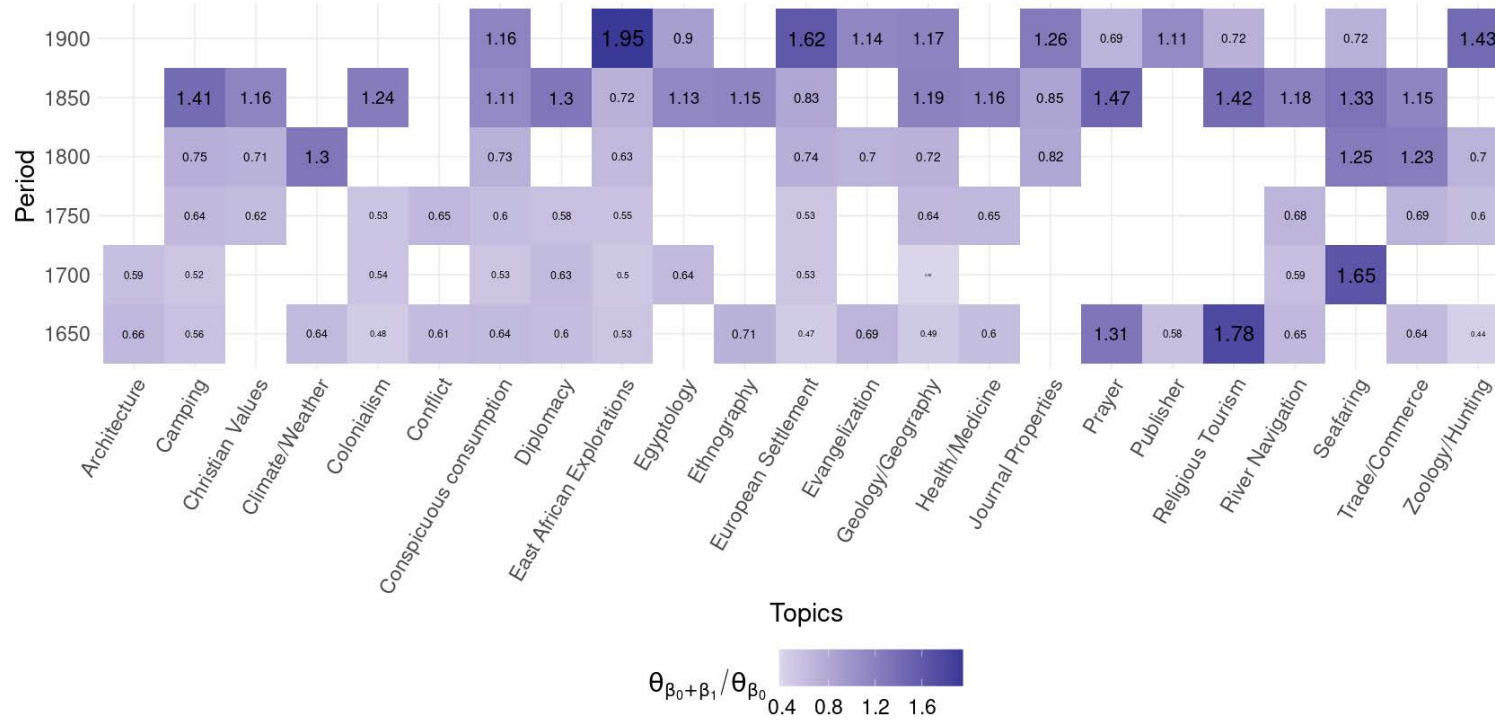


Figure 14: Results of the beta regression where proportion of topic, θ_d is modeled as a function of epoch in time. Excluded topic “Early Exploration” due to very large proportions of between 4.0 and 9.0 between 1650 and 1800.

I Word Clouds

beautiful goddess gateway entire modern remarkable finished
 means manne circular intended feet height hard hundred feet
 feet length thirty stand dimensions sculptured twenty times remain
 interesting block distance lines depth osiris court date nile
 columns sacred bottom depth osiris court date nile
 twelve dynasty interior sand face north diameter similar foot king
 water arch head place thebes hundred chambers west point
 inside is broken work time ancient granite roof opposite leading
 plateraised statue entrance length building in floor plan level
 position space remains walls east long front round light marble
 rubbish style half king high walls side chamber round body usual
 year called square found feet treat rock size deep pottery
 group thick back over figures temple small rock size deep pottery
 greek door ruins tomb part stone made hall open opening
 edifice colossa parts upper height pyramid large built entress solid pillar
 of degypt figure covered form egyptian base find wood
 limestone hand line passages site inscription feet high steps mummy
 sculpture handle hieroglyphics feet long steps close dead
 portico wide stood stands south kings standing middle laid
 floor original period outer area earth breadth perfect broad
 sarcophagus beneath construction principal masonry fallen he was supposed
 dedicated apparently single enclosure structure fine doorway
 evidently

(a) Architecture

clothes fatigue paid exhausted travels waiting midnight asleep noise
 heat tribe continue felt servant late foot sunrise difficulty sitting work
 spot forward children provisions making leaving drink visit stay
 load arrival breakfast baggage evening felt sunset chap fear
 wills thirst large began sheep arab country thing glad
 she the cattle milk make desert back caravan thing dark route
 reach pass afternoon road hour till good small slept life
 dinner carry gave evening found days find half
 stoptown thought journey people great or set taking tobacco
 meetyoung bread place made night village old distance lost
 things order chierest morning time left came usual mule
 cold fast rode rice upper fire camp ten water hours half ready full
 bring head ground brought party long food early coffee received
 load sand short horses sleep women return heard travel tired
 close start guide pool found marc obliged part cormoon
 manne meal house found marc obliged part cormoon
 friend called number travelling meat knew hands present finding
 moments health mounted supply hand appearance prepared
 immediately fresh wanted minutes saddle scarcely side encampment care
 half hour path anxious crowd miles departure search hiding trouble escort
 expected

(b) Camping

education peace receive self property respect jesus christ
 sacrifice desire frequently self property respect jesus christ
 manner reason practice subject small reading daughter
 hold light number knowledge fact common person times setish
 happy case chief generally nature money dead hand moral
 rest white boys slave thing called holy words find held daily
 backsold body yesu family called holy wive though books
 born girbospel make death love religion soul kind eam
 form days space poor woman good dread present believe voice
 free spirits girls power made children great worshipping pray
 head living vif faith wife people work house round understand
 sense call long heart time life women word truth master paid
 bride land custom true young things world country divine feeling
 orde brought old father years spirit child mind bibleature
 state heaven home full native husband dive prayer priest left
 face lives marriage missionary heard married earth hands class
 show language missionar heard married earth hands class
 services sacred character school prayers part christianity
 story teach prophet koran customs influence bring
 ceremony grave spiritual hope matter considered village grace
 africa superstition blessed savage hope teaching marry english

(c) Christian Values

influence moderate extreme season year higher find hazy
 covered extremely fallen are under lightning stars water found
 sun shine burning faller are under lightning stars water found
 ground heavens dark scarcely sudden intens horizon effect breath
 dul country shelter life times heavy rain fresh fall spring gradually
 greater felt october rainy season warm february january bright state
 round mist rest lightning hundred atmosphere water bega force
 feet and earth year calm light months sand south hail
 usual blow violent year calm light months sand south hail
 general fall high month winter night days cool continued
 oper blew lune clear wind cold evening shade east storm adding
 expose place guly till heavy rain season fell august pleasant
 half gale east winds heat weather rainy part state full vivid
 good fire hours morning storm time limited dust thick west
 moisture cloudy thermomet temperature summer north effect
 thing rise noon november cloud breeze snow december rain fell
 degree power afternoon pas generally nights slight nature quarter
 sere violence early sunset sunrise damp oppressive experience sleep
 beginning appearance period frequent drops seldom frequently
 flood tropical moment scene deep tempest height week
 average experience

(d) Climate/Weather

scenery ready rest village retired determined hartoum making
 occasion strength arrival driven whilst flank advancing twenty
 news courage point occupied direction round short
 move siege distance fall point occupied direction round short
 fled on action open campaign work retreat native move
 post guard till horses horse british mounted battery
 brave head fight farms line military artillery staff
 heard charge fell cavalry hundred position fighting fifty join
 good side front fire colonel office order found fifty join
 main miles ear captain force enemy french soldier fought
 lines firing back left troops battle house and began
 division field major army time great ground victory
 gallant strong night camp general party corps stores
 effect road orders made march wounded ordered chief
 make column regiment fort command shot close geturr
 defence down body place numbers escape advance small hands
 bush hill lieutenant advanced garrison infantry country
 face forces commanding ammunition brigade loss service rode
 death numbers heavy duty full guard hand gave company
 commander english cattle word engaged cannon transport
 considerable leaving success royal square lord hours movement
 coming alarm gordon rank report

(e) Conflict

beautifully stands band oriental carriage chap lined kitchen gallery
 shop finest wide procession king elegant fashion black
 dishes blue variety verandah gold silver hatched coloured
 opposite narrow chief ornaments brass comfortable seat
 half music beauty apartment family women occupied
 present queen french grand rich gardens taste market
 general light found town wood good dinners space curious
 splendid native windows hotel place high city make
 inside entre english gilt floor rooms sadies green walk
 pipes time long side small house built glass coffeesize
 yard piece beautiful gold large covered front filled close
 capital visit wooded room made prince kind form cool
 view full round door great walls made prince kind form cool
 residence rich feet table open square round white clean
 brought bar garden street number foot marble or left
 wall half style handsome people silk royal clay corner
 india dressed appearance public called furniture heat broad
 pretty excellent laid water principal generally
 interior country chair numerous entrance magnificent crowd
 spacious moorish lofty stood bamboo private furnished
 carpet paper courtyard manufacture usua thing japanese
 picturesque held business twenty order picture scarlet
 scene situated

(f) Conspicuous Consumption

respecting fact september command communication august hundred january demand british resident safety high doubt district reason position april inform taking appointed territory affairs interview expected instructions effect possession purpose writ occasion fear remain protection matter permission declared person reply gordon messenger hands brother request friendly heard intended subjects terms mohammed death majesty departure friends sudan bring good day letters visit power cattle peace presents major cambridge great give order arrival hewswritten meet agreed force pasha country received orders march object attack found french case king make back head honour capital gave place people letter british leave number delay till paid long consul governor account thought stay arms coast sultan made government returned friendship white conduct english expedition byssinia resident chap authority ordered brought part hope party informed receive journey mand answer captain report called large immediately message missiortroops sheikh report called information excellency intention knew money months years determined abysinnian assistance province late proceed obtain ready coming anxious theodore opportunity manner army reported october february considered stated presence attempt goods month desire june native things making grade prince principal interpreter

(g) Diplomacy

informed language mahomet belonging receive months governor piece principal fifty opinion heard betwix continued inhabited chief forty league prince general received commonly black lifewood making head general twelve return generally black counthither fell womer length north live twenty nature ands plenty company world years sound live twenty nature ands treight carry hath number parts brought fold port fall mouth find city people land hundreded defeat cafe wine thirty kind manner land hundreded defeat cafe fear leave goods fort time king water half french hold mile proper gave fame great make givrich court poor children rabia find ul gold made country account hand egypt drink quantity india cal place part good year here religion guinea north kingdom town trade travels reftaid loft china farther bright day high times high voyage ground nation fold iron earth obliged prince feet money point means safe wet taking immediately create here of foot money wet taking army abundance portugall thiopachap farms thought fame time notice true regard captain africa dead white began middle numbers danger journey europe famous contrary divers likewise grand algiers mountain open divers likewise

(h) Early Exploration

making jake albert kikuyu island friendly received hunting travel geographical banagas lieutenant major east african eastern africa stores means sudanes death difficulties plateau thombasa equator bring deal state heads protectorate steamer general southern lines small brought equatorial bantu ganyika annava province place goods food return white south visit forest company burtor number star left days great ivory station west sultan leaving position life falls people great favor station west sultan leaving european fort caravan centra stanley chief british kampala port pasha grant congo african ile mission point give north upper arab coast uganda march tribe king heard bar present arab coast uganda march tribe king heard etter view good time expedition africa north told count apri banana load living stone zanzibar work large home june leave make years central africa victoria chapter made easterr teach egyptian here east africa journey long back rived district round feet stairs miles found regions news captain baker village round wilson months soma transport region party land exploration english language carry swahili index territory general grant ammunition relief kilimanjaro dark continent stanley pool called stanley fall wenzori british east coast question columbus basin story slave free cattle salt

(i) East African Explorations

celebrate east meaning required subject inscription usual kings stance festival lines ethiopia appearance population sufficient plural passage turkish considerab sacred expens population sufficient plural grand frequently case banks money mode modern nubian object labour hundred give times thousand dragoman emperor boat verb purpose ossetta european capital description king attention public large english lower means slower egypt account fact arrival foreign french days part ancient life found ling route doubt sall similar hediva place hebes made upper paid home interest custom greater make place canal egyptian traveller chapter chief remain person order china cairo nubia built europe part river royal common aras suez egypt time present brought antiquity principal word alexandria Nile time present brought antiquity thing emphasis called chinese greatness delta atarack water works power period general desert embassy ver country visit tomb coptic date generally years viceroy anton upper egypt arabic scarcely cost palace temple embassy china number perform long scene word voyage journey egypt nubia character antiquities numerous view inundation making return inductor ruins people kind state bank idea respect manner considered suz canal herodotus good curious singular eastern month extraordinary small interesting degree family banks Nile presents construction island language arabian intended situation amount manners construction island language arabian intended

(j) Egyptology

distinct brought interior horn curved hindler rounded closely good clothing trousers garment beauty black white ornament surface centre hanging peculiar arab flat handsome close complexion beautiful line shirt strings specimen quill coloured beard called brass baseth male common negro throat features bird leather from arms male female found edge time bright nose species heads covered appearance plate pointed make pair iron skin feet feather length hand kind dirty loosh high half cloth large blue womer worm silk number pretty hands wore yellow long head dark pieces shape hung strong parts beads small white row part young foot coat copper thick light back half black body middle outer wing figure colour round naked chin bare figure teeth dress neck wear tail side haked country similar broad upper made face legs form size wide gold traigh ring deep broad upper made face legs form size wide gold traigh scarlet urban generally green shoulders lower ale pointing string band wings bill grey cotton dressed ears pale mouth silver slightly waist narrow ornaments people costum sort race belly slender tied shoulder scales marked general africa covering wood ground fashion soft yellow fest reddish woman drawn frequently round neck yellow purple feather orange frequently coarsedies dance distinguished

(k) Ethnography

neighbourhood store night zululand open englishman excellent years south living goods kind drawn amount heavy vehicle black ground return taking team difficulty karro bought quartz generally north ride leaving hundred miles knew wheels hill elizabeth capital port elizabeth diamonds free thing money drive british hard stock transport travelling gate things hundred distance left brought hottentot five drift visit carriage kraal pretoria district land port high give twenty life fields dutch great transvaal land port high give lost half line colony gold horses railway white fact worth kaffir eatape tow cape town kaffir called table system car place found south africa days cost drive young sheep back on road miles oxen boer part making means make south africa natal years native present horse wagon time cattle mine house iron african horse wagon time cattle mine house iron mining wagon side arm zulu good velvet fine large journey diamond english fort coal john farmer fold small roads chapter company travel doubt months england south african zambesi durban game number cape colony victoria station thousand route party business trip australia trek zambesi durban rough orange bush fest carry round week labour matter spent white fest carry round people stores heard thought hours salisbury experience thirty future spot required johannesburg paid hunting

(l) European Settlement

coming minister india miles father engaged south building excellent night short african march september voyage heard back thursday leaving twelve wrote november tolbourn gentlemans fell lady january christian present make hotel months catholic kindly brother charged death hear french friend lord board africa young saturday wednesday port sierra james school house leon late land afterwards services bishop home return happy messrs chapter england home return party states george placetime church great friday governor july dear visit mission town cape long friday pleasant thomas john work made called charles united states evening good missionary english prim family america work days sunday years native gave children leave station william captain london appointed pete early arrival letters hope american week august society miss read hope people kind dinner company united sierra leon chapel found mail city marysland departure services congregation found mail city feed interesting find sabbath october december tuesday brought kindness month pleasure number february account state west reading liverpool missionary society robert protestant care country consul castle write brethren journal give office thought smith johnson

(m) Evangelization

remains possession christianity earlier results mention barbery details tradition distinguished preface romans palestine turkey throne manner national civilization easter earliest customs complete record influence literature regard states museum translation description mind hebrew founded royal introduction public foreign travel point state family christian full land arabic author origin notes natural doubt xristian india english france works remarkable ideavolume times found subject general original days learned long greek found subject general original days text true foreign empire french period word note page means important year egyptian ancient early kingdom egypt give num letters king time history account people paris list algiers religion present egypt country people paris list truth england king world great work modern reader view morocco rom language century made facts yria arts european knowledge power european interest rome emperor nation place information east read a death italy called question narrative religious quest asia father writing number study opinion science importance books evidence prince progress political existence bible valuable position british make beginning abyssinia thousand common native attention capital connected italian established german constantinople points geography system collection traveller points geography system appendix

(n) Fertile Crescent

lying gorge wells perpendicular bas information portion arab remarkable general bare elevated reach beauty considerable dead cultivate picturesquescattered districts descended limestone elevation bottom gulftousand wall fine heights lava situated lofty vegetation hundred mountain masses hours round half ascent lower surface distant form clear broken level tall place distance base rise find length chain ridge land view hills great angle narrow hour remains small miles rock plains south ruins rising nature town western west rocks valley side high sand numerous spring higher village par mountain sinaisnow stood beautiful deep water road feeh north rocky path wide green made earth stone foot mount east summit left route mass upper line ancient country long covered time was natural sandy slope eastern southern ground plain suggested atlas walls northern region spot city direction granite wild close volcanic edge avine site sandstone present peak broad descent barren sight passing visible rose blue white position hundred feet beneath wilderness running opposite built gradually leaving precipitous journey farther dark winding verdure aspect traveller times neighbourhood eastward

(o) Geology/Geography

deal experienced feeling terrible attention physician pulse ive prevent practice fatigue consequence required evening temperature result appearance dangerous repeated young manner leave received dead worse affected short white animal cholera present heat africa thing thought called matter number work half goat fact generally frequently applied morning continued suffer period head effect poison morning continued suffer common treatment added make condition mouth form hands skin effects felt left high patient gave native daily severe found doctor health lift fatal person month usual drinking cure pain sick water good long bottle mind drunk diet means life great fever days body taking narc dose brought death time disease drink kind hand year feel told care state made medicine case attack give obliged res health place food medicine case attack give obliged painful large people suffering part poor sickness quantity bath milk back months till weak climate cold strength years warm nature sentry country return plague stomach began order recovery operation hospital chap spirits unable quinine general change immediately dying stage pains relief slight exposure tongue mortality frequent stage pains relief account removed greatly face sleep sufficient exposed

(p) Health/Medicine

sufficient mind appointed terms tribe west individual proposed none authorities reason establishment account control civilization republic portion possession special considered brought progress manner measure committee considerable cheuanaland form desire things late established foreign duty peace respect force protection labour south africa fact states territory circumstances east office population governor order means society success nature town question english good public power authority district secretary french influence british great times subject found duties view year trade people state made lav dong hands importance high free government chief character rights french hope policy part country native place interests life effect affairs work colony paid africa land president received services system years cape present number money doubt give england case dutch south position india united european property slavery justice company military powers matter transvaal free state missionary orange purpose regard administration large important political amount letter treaty greater report knowledge support britain plan measure person conduct advantage slave trade commercial extent matters cape colony land necessity difficulties feeling revenue personal degree independent education interior action called proper charles

(q) Imperialism

subject satisfaction repeated length side dinner hast journey husband hearings slave obliged opportunity world making wanted talk understand full show countenance truth desired sitting business hour room presence matter bring reply feeling answer kind find servant pleased spoke called tear season service royal country began back home manner coming paid state love years left brother hands friend place family lead peace master words give lord household visit evening gum pray palace fear words give lord household visit evening gum held things young hand king made sword felt hope doubt child queen lady make told good deer father days rest desire speech ordered mind death thou thought thing court favour taking speak thee gave people order meet cried money mind death thou thought thing court favour daughter ready voice knew return present brought face true read bar prince knew return present brought face true read feeling immediately moment conversation mother doot continued dear pleasure might woman majesty account happy cal stay sister receive english honour care usual gentleman interpret company kindness short attention informed dost language story party children means respect silence case language party children means respect silence case language conduct health meeting custom afraid arms arrival trouble white

(r) Prayer

eastern quarters moslem ride patriarch interior early famous mohammed women ewish ride olives syrian early fountains voun palestine brough prophet chapter door occupied john priest miles distance principal jaffa governor david priester miles distance principal jaffa governor david hand country appearance arab garden modern citadel walk world residence sepulchre left sacred made item mousa usordemight bridge half palace house and village spot hundred east stands standing jesu quarter years mount ruins ancient stands standing ruined saint house ruins ancient stands standing mass situated part arg holy place sultan remains covered plain space aid ord street time vivaldi city built christian temple tunisizion algeria round french high greek town church walls site foot horses public rich marble side mosque gate called tomb altar tradition dome scene close tower convent small stone visit one enter cave latin good building turkish gates chapel road step mary centre population beautiful fine hill gardens found view grand held rom holy land days full present cross view grand constantinople number water morocco cairo narrow gulf mile strong bethelem fortress times entrance moorish fine pilgrim age rock dead make magnificent pass times entrance moorish fine pilgrim age rock dead egypt holy city nazareth poor description journey received grotto beneath mountain

(s) Religious Tourism

upper times westward wooded half mile shallow neighbourhood number direction heard situation hands french frigate cast harbor coming view stood thing schooner mat provisions twelve bound appearance return longitude pegan making officer craft brought continued hundred high tide immediately traits head fine hour till weather north light in day reef strong evening harbour sur steamer left east danger though fleet half crew isle found night can order early reach fresh side vessel captain miles cabin winds east world length part anchorage island boat made sight back fish taking current cape time board wind great close full tons past steam round port shore ship coast hours short fair bridge place days water voyage land long make ready leaving calling morning lock sail deck point rocks flag fell distant shore morning good beach small landed post calm leave noon passage good beach small landed post calm hope pilot distance breeze latitude cargo landing clear safe sand south gave large ocean west english post calm town afternoon obliged bottom anchorage minutes arrival pass heavy channel rock dark twenty gale stern running mile company alongside departure dangerous passing running mile beautiful eleven watch overboard deep account commodore work usual opposite watch determined deep account main rope struck

(t) River Navigation

waters wreck board ship mouth spanish bombay sailor gibraltar direction heard situation hands french frigate cast harbor number view stood thing schooner mat provisions twelve coming appearance return longitude pegan making officer bound appearance return longitude pegan making officer craft brought continued hundred high tide immediately traits head fine hour till weather north light in day reef strong evening harbour sur steamer left east danger though fleet half crew isle found night can order early reach fresh side vessel captain miles cabin winds east world length part anchorage island boat made sight back fish taking current cape time board wind great close full tons past steam round port shore ship coast hours short fair bridge place days water voyage land long make ready leaving calling morning lock sail deck point rocks flag fell distant shore morning good beach small landed post calm leave noon passage good beach small landed post calm hope pilot distance breeze latitude cargo landing clear safe sand south gave large ocean west english post calm town afternoon obliged bottom anchorage minutes arrival pass heavy channel rock dark twenty gale stern running mile company alongside departure dangerous passing running mile beautiful eleven watch overboard deep account commodore work usual opposite watch determined deep account main rope struck

(u) Seafaring

dried collected industry extremely family eaten bark similar proportion height peculiar taste character angola niger remarkable western africa west coast frequently england vegetation labour central region considered eastern iron natural tropical chiefly growing america sierra leone nature continent abundant guinea brought variety wheat white gold coast southern commerce abundance work upper fish south indian india extent cultivation wood district growth size native interior african cotton palm live sold western land soil fruit part cultivate make french european cape island found goods supply place sweeter form leone cattle gold great africa large year race vegetable root state grow king plant country species made quantities world english rice called trade tree small water negro lower excellent article produce corn parts food quality ivory roots means common produce corn leaves wild feed find rich time portuguese years climate considerable wine numerous west africa sierra high general ground fact fine nadagascar population east europe salt hundred account greater tobacco scale principal sugar regions coffee season thousand gardens principal senegal valuable coffee extensive growth imported sufficient fertile portion important numbers lands agriculture amount fertile portion important numbers northern purposes goods description appearance times equal twenty france manufacture helen ansed distribution easily

(v) Trade/Commerce

native forward piece follow curious people occasionally appearance straight africa minutes smoke chance twenty sport escape bullet creature struck frequently lumber shoulder scene shooting direction standing snake shoot africa nest woods leopard cattle kind brought covered food whist pretty kill hunting moment white flesh sight dark find home bull buffalo dogs night dead back camp find generally blood meat bird lion great found caught sharp tail length size open game wild side time heard haphunt lying fine head trees hot long fire high animal sand give gave path feet animal times turn sand part jungle full forest ground grass made left wood called fell till close found large water herd short began trunk noise beast elephant bush small leaves numbers slowly rifle green young good distance leaves numbers broken life hand horse black suddenly thick skin rest dense face huge body legs country deep rhinoceros heads wounded hard place making beautiful high beneath plain approach hole hundred morning party common light beneath shaden numerous immediate lands coming strange at front heavy hippopotamus strong stream poor walk edge

(w) Zoology/Hunting

Oxford Economic and Social History Working Papers

are edited by

Mattia Bertazzini (Nuffield College, Oxford, OX1 1NF)

Marco Molteni (Pembroke College, Oxford, OX1 1DW)