
Benjamin Martin ¹, Fredrik Mohammadi Norén ²

¹ Uppsala University, Box 629 751 26 UPPSALA
² Malmö University, Box 50500 202 50 Malmö

Abstract
This study uses digital text analysis, focusing on LDA topic modeling, to conduct a historical investigation of the relationship between the concepts of nature and culture found in the pages of the official magazine of the United Nations Educational, Scientific and Cultural Organization, The UNESCO Courier, between 1948 and 2020. The relationship between the concepts of nature and culture has historically been at the core of concerns about the environment and sustainability; Courier offers a means of charting a global conversation on these concepts. After presenting the corpus and our methods, the paper documents three approaches to LDA topic modeling that we have tested, through which we seek to make topic modeling useful for the field of conceptual history. Our empirical findings suggest that the concepts of nature and culture have come to be increasingly close over the course of the last six decades, while the stakes of the very distinction between the concepts have changed radically. Our methodological tests support the argument that topic modeling can be a valuable tool for conceptual history, albeit one that must be handled with care.

Keywords
Conceptual history, Global history, digital text analysis, topic modeling, UNESCO

1. Introduction

UNESCO is the United Nations organization for education, science and culture — not, in the first instance, for the environment or nature. But the organization has a long history of concerning itself with issues related to the natural environment. UNESCO’s monthly magazine Courier reveals a striking level of interest in such issues. Since the magazine’s foundation in 1948, Courier has featured articles that documented diverse ways that the world’s peoples live in their natural environments, explained how particular cultures shaped landscapes, presented breakthroughs in scientific knowledge about nature, celebrated efforts to preserve humanity’s “natural heritage”, and, more recently, discussed the role of human activity in changing the Earth’s climate [1]. At the heart of each of these topics was a set of fundamental questions about the relationship between nature and culture.

In the history of thinking about sustainability, the environment and related issues, the shifting and competing understandings of the relationship between nature and culture constitute a fundamental element. Indeed, the nature-culture dichotomy is a classic theme in the history of concepts. After all, as historians have shown, the concepts of culture and nature have always existed in relation to one another: defining a sphere of human autonomy and self-fashioning (culture) requires an opposing category encompassing that which humans did not create and which they struggle to control (nature), and vice versa [2]. The discussions of nature in Courier, a journal devoted to the themes of education, science and culture, offers an interesting source for exploring the history of that dichotomy in a dramatic phase in world history, during which the natural environment came to be a topic of international concern [3].
The presence of nature-related themes in *Courier* is of particular interest, moreover, because of the publication’s global character. Founded to “promote UNESCO’s ideals, maintain a platform for the dialogue between cultures and provide a forum for international debate,” *Courier* had uniquely global aspirations and reach. At its high point in the 1970s and 1980s it featured articles from prominent intellectuals across the globe published in 35 languages with an overall distribution of over 1.5 million copies, and was available on both sides of the Iron Curtain [4]. This magazine was recently digitized and made available by UNESCO (en.unesco.org/courier/archives). Working with developers at Humlab (Umeå University), we are curating this archive into a machine-readable corpus suitable for digital text analysis.

In this paper, we use *Courier* to follow a global conversation on the relationship between nature and culture. Given the magazine’s focus on cultural matters, one means of charting that relationship is by focusing on the uses of the concept of culture therein. We do this by deploying tools of digital text analysis, in particular LDA topic modeling, in combination with close reading, to locate contexts in which the concept appeared in the magazine, measure how connections between these contexts changed over time, and identify novel ways to chart changes in the way the nature-culture relationship was articulated.

On the basis of these three approaches to using topic modeling we make a set of empirical findings and a methodological argument. In empirical terms, our preliminary findings allow us to identify thematic contexts in which nature was discussed in *Courier*, as well as to rank those contexts in terms of their relative strength in the corpus. They also reveal a fundamental debate in the publication over how to think about the concept of “nature” – a debate that advanced through arguments about what the relationship was (or should be) between nature and culture. Regarding methodology, we argue that topic modeling can serve as a useful tool for conceptual history, provided one is careful about what each computer-generated “topic” is, and is not.

### 2. Corpus and Method

The source for this study is the text printed in *Courier* from the magazine’s foundation in 1948 until 2020. The print run was quite consistent for most of that time: the magazine was published monthly from 1948 until 2002, at which point *Courier* decreased the number of issues per year. After 2002, the publication rate changed often, and ceased entirely from 2013 to 2017, at which point it was relaunched at a rate of four issues per year. We downloaded the PDFs and used the open-source Tesseract OCR Engine to make the text machine readable. While not perfect, the OCR quality is good. Our own test of samples from the whole corpus generated an error rate of only 0.7 %. The total *Courier* corpus, from 1948 to 2020, consists of some 13 million tokens.

The most significant feature of this corpus is that, in contrast to large, general-language corpora (such as Google Books or Eighteenth-Century Books Online), ours is a focused, historically specific, and carefully curated text corpus. It is more specific in the themes it addresses than many other digitized publications, like daily newspapers, and it is linked in a specific way to a particular social agent: not just a publisher, but an international organization. This allows us to pose and hopefully answer questions of a sharper kind than those one can study through a broad, generic corpus, while also allowing us to link those questions to a social reality beyond the words we find in *Courier* – namely, the history of UNESCO as an institution. Moreover, and most importantly, insofar as *Courier* sought to give voice to representatives of UNESCO’s member states, this is a corpus that gives us access to something approaching a global conversation – something that very few other bodies of text can match.

Our method in this investigation focuses on LDA topic modeling. This is a probabilistic method suitable for structuring a large and diverse text collection. Based on word distributions in documents, the model assigns each word a probability value and structures them into topics, and packages the underlying documents into blends of topics. Topic modeling builds on the principle that a word can be part of several topics, and that all topics are distributed in every document, but with different degrees of probability, and sometimes with very low values. There are different topic modeling algorithms, and this article uses the popular Latent Dirichlet Allocation, as implemented in Mallet [5]. The researcher decides on the number of topics to include, but after that, the model works unsupervised – inductively,
one could argue – to identify “topics”: top lists of words that have a higher probability of occurring close to one another in the different documents.

The researcher can then interpret and label these “topics” to capture the theme that each collection of words seems to reflect – often by moving back and forth between the statistical results (the lists of words) and passages from the documents themselves, in order to identify the sense of these topics as clearly as possible. Usually, several topics are more or less easily interpretable, while others are quite opaque. In the best case, allowing LDA to identify these topics offers “a method for analyzing texts...that is substantively quicker, more efficient and more objective than traditional methods of content analysis in the social and cultural sciences” [6]. At the same time, care is called for in handling these “topics”, particularly for historical research. Each topic identifies real features of the text, identifying collections of words that may (in the best cases) correspond to what linguists (and some conceptual historians) call a “semantic cluster”. But computer-generated topics cannot necessarily be understood as a “theme” or “discourse” in the sense in which these terms are use in the humanities and social sciences, much less as a “concept”. If we are clear about these limits, however, the ability of the method to offer a fresh take on a large corpus has advantages for the history of concepts that we aim to explore here.

Before producing the topic models of Courier, some corpus preparations were made. We performed a part-of-speech tagging of the corpus and excluded numerals, delimiters, and all tokens that occur less than five times in total. A single page was chosen as the “document level” for creating the topic model.

Our text curating will soon enable us to analyze at the level of articles, instead. For now, we used UNESCO’s own index of Courier articles to automatically identify and extract the content of each issue from the first page on which an indexed article appears to the last. This excludes non-article content such as publication information (the publisher's imprint or “masthead”), tables of contents, editorials, and letters to the editor. Courier is a magazine free from traditional advertisements, and thus mainly consists of editorial content, including articles, photos with captions, and sometimes tables and graphs.

Having computed several different models, we selected a model of 200 topics (calculated on the basis of the corpus as divided up into single pages). We interpreted these topics manually and assigned each one a thematic label. A Jupyter notebook environment was set up to explore the topics through different tools, including topic word distribution, topic over time, and topic networks.

To use topic modeling for this historical investigation, we divide Courier’s print-run into sub-periods, so as to be able to observe change over time. Inspired by the periodization used in Cholé Maurel’s history of UNESCO, which identifies periods corresponding to the tenures of the Directors General, we divide up the corpus into the following three blocs [7]:

- Phase 1: 1945–1961: From Huxley to Veronese (16 years)
- Phase 2: 1962–1986: Maheu and M’bow (24 years)
- Phase 3: 1987–2019: From Zaragoza to Azoulay (28 years)

These periods correspond moreover to broad changes in UNESCO’s history. It was in the early 1960s that the organization’s membership expanded with the accession of many postcolonial states; the mid-1980s is widely seen as having marked a political-ideological transition, occasioned not least by the withdrawal from UNESCO at that time of the United States, Great Britain, and Singapore [8].

The methodological core of this article is an effort to use LDA topic modeling to explore the history of certain concepts. We seek thereby to contribute in particular to the recent research in transnational and global conceptual history, as well as to the scholarship in what can be called digital conceptual history [9]. Work in this emerging field has explored a variety of methods, including word-trend and collocate analysis, word embeddings (or vector-space models), as well as network analysis of co-occurrence data [10]. What can topic modeling offer to these efforts? It is by now relatively uncontroversial to claim that topic modeling, by providing “a way for researchers to obtain reasonable automated content coding of large text corpora,” can offer a stimulating means of exploring a large body of text, insofar as the algorithm often suggests perspectives and connections that would otherwise be missed [11]. But this is still essentially a means of determining what is in the corpus. Here we ask whether topic modeling can help us understand not just what was discussed in Courier, but how it was discussed—and how that discussion changed over time.

In what follows we use topic modeling in three ways:
1. We apply word searches to our topic model as a means of identifying thematic contexts in which the concepts under examination appeared, and in order to explore the relative strength of those contexts in the corpus.

2. We apply network analysis to our topic model in order to identify relationships among contexts in which the concepts appeared.

3. We use the topic model to identify groups of pages in the magazine in which a chosen topic (which contains our target concept words) was particularly strong in a selected time period, for manual reading and analysis.

3. What did *Courier* talk about when it talked about nature?

To examine discussions of nature in *Courier* we need first to locate these. To do that, we identified all the topics (of the 200 generated by the algorithm) in which the words ‘nature’ or ‘natural’ appear among the top 50 words. One could of course simply search for all appearances of the words ‘nature’ and ‘natural’ in the corpus. The virtue of instead following these words’ appearance in a topic model is that doing so leads us immediately to a set of thematic foci, which one would otherwise need to build up from scratch. The themes suggested by the topic modeling output are, of course, not themes that we designed or selected in advance. Topic modeling serves us, then, as a data-driven means of identifying the contexts within which nature was discussed in *Courier* in a manner that is able to generate discovery and surprise from the very beginning.

Proceeding in this manner identifies twenty unique topics, about which we can make several observations. Each “topic” consists, of course, simply of a ranked list of words. Happily, many of these topics are readily interpretable and seem easy to name. For example, topic 106 (“park”, “species”, “conservation”, “biosphere”, “national”, “reserve”, “natural”, “reserves”, “nature”, “world”), clearly has to do with national parks and other types of nature reserves, and is thus labeled “national parks”. Topic 50 (“earthquake”, “disaster”, “earthquakes”, “damage”, “floods”, “warning”, “disasters”, “tsunami”, “natural”, “caused”) addresses natural disasters, and is labeled the same. Others are relatively clear, but require more context to understand. Topic 193 (“human”, “man”, “nature”, “life”, “biological”, “living”) seems to refer to the human species as part of “nature”, but in an ambiguous sense. In this case, we labeled this topic “human species”.

<table>
<thead>
<tr>
<th>Topic number</th>
<th>Topic label</th>
<th>First 10 words</th>
</tr>
</thead>
<tbody>
<tr>
<td>157</td>
<td>knowledge</td>
<td>systems system based terms nature concept role process forms specific individual</td>
</tr>
<tr>
<td>198</td>
<td>experience</td>
<td>s world life time reality human nature form sense man</td>
</tr>
<tr>
<td>97</td>
<td>world civilization</td>
<td>world man great human men history today civilization mankind time</td>
</tr>
<tr>
<td>140</td>
<td>landscape</td>
<td>region river area land mountain mountains south north great valley</td>
</tr>
<tr>
<td>23</td>
<td>research</td>
<td>study research information studies data scientific work results made carried</td>
</tr>
</tbody>
</table>

2 Eleven topics feature the word “nature” and another fifteen feature the word “natural”; six of these overlap with the first list.
Second, some of the topics are collections of words that are harder to make sense of as a “topic” in the sense of a semantic theme. These suggest rather a type of language, or perhaps genre; a collection of terms that appear together because they are used to articulate a view of or attitude toward a given subject matter. We call these “genre” topics. Topic 157, for example, begins with the words “system” “based” “terms” “nature” “concept” “role” “process” “forms” “specific” “individual”. These terms suggest a metadiscourse about our knowledge of some other topic or topics. We have chosen to call this topic “knowledge systems”, while acknowledging that it is hard to tell what the meaning of this topic could be, in the sense of a semantic cluster. (“Nature” here might for example more often signify “the nature of something”, rather than the natural environment.) Topic 140 (“region” “river” “area” “land” “mountain” “mountains” “south” “north” “great” “valley”), similarly, collects language that might be used to describe landscapes of various types; we label the topic “landscape descriptions”.

<table>
<thead>
<tr>
<th>#</th>
<th>Topic</th>
<th>Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>36</td>
<td>science</td>
<td>science scientific research scientists sciences knowledge technology scientist natural</td>
</tr>
<tr>
<td>49</td>
<td>water</td>
<td>water river rivers irrigation dam supply waters dams fresh sea</td>
</tr>
<tr>
<td>156</td>
<td>heritage</td>
<td>heritage cultural world monuments sites unesco list site restoration conservation</td>
</tr>
<tr>
<td>193</td>
<td>human species</td>
<td>human man nature life natural species environment 's biological living</td>
</tr>
<tr>
<td>163</td>
<td>philosophy</td>
<td>philosophy 's philosopher avicenna works thought work knowledge philosophers philosophical</td>
</tr>
<tr>
<td>106</td>
<td>national parks</td>
<td>park species conservation biosphere national reserve natural reserves nature world</td>
</tr>
<tr>
<td>58</td>
<td>trees</td>
<td>tree trees red green white flowers colour leaves water garden</td>
</tr>
<tr>
<td>46</td>
<td>plants</td>
<td>plants plant cells chemical organisms substances cell bacteria micro process</td>
</tr>
<tr>
<td>104</td>
<td>climate change</td>
<td>climate environmental change global environment planet carbon world emissions earth</td>
</tr>
<tr>
<td>116</td>
<td>forests</td>
<td>forest forests trees soil tropical land tree erosion wood environment</td>
</tr>
<tr>
<td>50</td>
<td>natural disasters</td>
<td>earthquake disaster earthquakes damage floods warning disasters tsunami natural caused</td>
</tr>
<tr>
<td>20</td>
<td>natural resources</td>
<td>oil iron mining salt copper mineral 's deposits gold coal</td>
</tr>
<tr>
<td>91</td>
<td>indigenous peoples</td>
<td>peoples indians indigenous primitive tribes indian people tribe tribal men anthropology civilization</td>
</tr>
<tr>
<td>110</td>
<td>sounds</td>
<td>sound noise sounds ear waves hearing vibrations heard vibration noises</td>
</tr>
<tr>
<td>102</td>
<td>Darwin/evolution</td>
<td>darwin species whale whales islands galapagos evolution whaling charles natural</td>
</tr>
</tbody>
</table>

Table 1. Nature/natural topics in *Courier* (out of a 200-topic model), in order of strength/weight in the corpus as a whole.
Third, we can also observe that the results we get from topic modeling reflect particular features of *Courier* as a publication. For example, most issues of *Courier* were partly or entirely devoted to a particular theme, generally featured on the cover. This feature of the magazine means that a theme that may be rather infrequent in the corpus as a whole is nonetheless distinct in a small number of pages of the magazine (in the relevant special issue or two); this in turn raises the likelihood that such themes will be identified by the topic modeling algorithm as distinct topics. Examples of this are topic 102, labeled “Darwin/evolution” – the subject of a special issue in May 1982 commemorating the 100th anniversary of the scientist’s death – and topic 91, regarding indigenous peoples (and labeled as the same). Quite weak in the journal as a whole, this latter topic spikes in likelihood in particular years (as we can see using a tool for measuring topic trends over time). Charting topic 91 then with a tool designed to identify particular pages of *Courier* in which a topic appears most strongly leads us to the thematic issue of summer 1954 entitled, “Last Frontiers of Civilization,” which explored “the problem of the world’s primitive peoples” [12].

Fourth, we can use the quantitative data from which the topic model is constructed to measure the relative prevalence of these topics in *Courier*. The model’s measure of the likelihood of these topic’s appearance in the corpus as a whole offers us a quantitative means of ascertaining these topics’ relative strength. This shows that some of the broad “genre” topics in which “nature” appears – such as knowledge systems (157), experience (198), or landscape descriptions (140) – are among the very strongest (most likely) topics of all.3 Because they are so generic, these “topics” appear with many topics (in a narrower, semantic sense of the word) in the corpus. For that reason, they are also of limited utility for our conceptual-historical investigation. Once we exclude these, we can use the same method to measure the relative weight of the remaining topics. Doing this is particularly illuminating if we apply it to the topics as grouped into categories.

We observe namely that the topics – after excluding the “genre” topics – can be divided into three groups. First, we find a set of topics in which nature appears as the subject matter of scientific research, including (in order of relative strength in the corpus): research (23), science (36), human species (193), and Darwin/evolution (102). Second, we see a group of topics in which nature is the object of protection or preservation, including heritage (156), national parks (106), and forests (116). Third, several topics address ways humans live in (and with) nature, in topics such as climate change (104), natural disasters (50), natural resources (20), and indigenous peoples (91). A few topics do not fit this scheme and seem indeed largely irrelevant to our investigation: topic 163 (philosophy), for example, includes the word “nature” but only towards the end of the list of its first fifty words. The distinctions among these groups are of course not water-tight; some topics, including 193 (human species) and 104 (climate change) are intriguing precisely for the way they appear to straddle these categories. It is nonetheless noteworthy that data on topic “weight” shows that the topics in which nature is discussed in the context of scientific research are, as a group, the strongest of the three categories. Those related to preservation come second. The topics we categorize as related to living in nature are weakest.

On the basis of this relatively simple approach, then, we find that *Courier* discussed nature most strongly in terms of science (one of UNESCO’s focus areas) and preservation (the heart of UNESCO’s commitment to the world’s “cultural and natural heritage,” which is of course the organization’s most famous program). Discussions of nature in the context of “culture”—either in the sense of the arts, or in the sense of particular ways of life—seem by contrast to have been less prominent. But the third group of topics, while “weakest”, is also the most difficult to define and, for our purposes, most interesting. So, likewise, are those topics which are harder to place firmly in one of the three categories.

4. Wider contexts of the nature concept through topic networks

Another way to deepen our understanding of the themes surrounding “nature” and “natural” in *Courier* using topic modeling is to study how the topics containing these two terms interacted with other topics in the 200-topic model. Since topic modeling builds on the principle that every topic is represented in every document, although in some cases to a minimal degree, it is possible to construct

---

3 Four of these topics (157, 198, 97, and 140) are among the 20 most likely in the entire corpus. The list of all topics from our 200-topic model, ranked by “score”, is online at: https://docs.google.com/spreadsheets/d/1-P59S_hvHnXTxsrnmDKHECmmdW41Bp/edit?usp=sharing&ouid=101084557857117261252&rtpof=true&sd=true
and visualize networks of co-occurring topics, based on topics that are present on the same magazine page over a given threshold. If studying a single topic says something about that topic’s inherent characteristics, examining its broader contexts, or topical associations, can expand our understanding of how, in this case, ideas of nature were envisioned in *Courier*. If a topic about climate change, for example, were to co-occur with a topic about pollution it would generate different connotations than if the climate change topic were to co-occur with one about heritage or art.

In order to study how topics’ co-occurrences have changed over time, we compare topic networks for each of our three time periods (1948–1961, 1962–1986 and 1987–2020). To count as a co-occurrence, two topics need to be represented on a *Courier* page at a weight minimum of 0.1 and must co-occur on at least ten pages in the first period and twenty pages in two later periods (in order to take into account the different sizes of the three sub-corpora). Topics that perform below the weight document-score thresholds are thus excluded from the networks. These settings are designed to reveal a network of topic ties that are both representative and strong. For visualization of co-occurring topics, we used the network tool Gephi. The layout algorithm Force Atlas 2 was employed to model the three networks. The sizes of nodes (topic labels) are based on the sum of all individual connections to a topic. The thickness of edges depends on how many documents two topics share. The results are displayed in Figures 1–3, in which topics that contain either “nature” or “natural” (among that topic’s first fifty words) are marked with red circles.

![Networks of co-occurring topics for the period 1948–1961. Weight co-occurrence score: 0.1, shared document score: 10. Nature-oriented topics are marked with red circles.](image)

**Figure 1.** Networks of co-occurring topics for the period 1948–1961. Weight co-occurrence score: 0.1, shared document score: 10. Nature-oriented topics are marked with red circles.
Figure 2. Networks of co-occurring topics for the period 1962–1986. Weight co-occurrence score: 0.1, shared document score: 20. Nature-oriented topics are marked with red circles.
Figure 3. Networks of co-occurring topics for the period 1987–2020. Weight co-occurrence score: 0.1, shared document score: 20. Nature-oriented topics are marked with red circles.

Studying the three networks, some general trends emerge. In each period, a few topics dominate and tie the network together. These include, some of the broad “genre” topics like topic 19 (problems (generic)), topic 157 (knowledge systems), and topic 16 (development). These topics are, as we have seen, among the strongest (by score ranking) in the entire the 200-topic model. Another observation about the network structure concerns how topics are clustered together. The earliest period’s network (figure 1) divides into two thematic clusters: one (on the left of the visualization) related to UNESCO itself, including its institutions, programs and two of its three core fields (education and science), and another (on the right) having to do with historical and contemporary cultural and religious expressions. Completely separated in the first period, these two clusters remain essentially distinct in the second period – they are linked only by topic 198 (experience), a “genre” topic that includes terms like “world” “life”, “time”, “reality”, “human” and “nature”. In the third period, however, it is difficult to discern clear borders between clusters. This seems like a significant change. What might account for it – a change in the content of Courier? An integration of the kinds of generic language use to discuss different topics? A change in the understanding of the relationship between culture and social life? – is impossible to determine from these networks alone. But the change itself raises interesting questions.

Several of the nature-related topics we identified through our first approach do not appear in any of these three networks of co-occurring topics. This does not necessarily mean that, for instance, topic 20 (natural resources), topic 50 (natural disasters), topic 91 (indigenous peoples), topic 116 (forests), or topic 156 (heritage) are not strong themes as such. It simply means that these topics are found to be less likely to co-occur with other Courier topics (given the network settings we applied). Indeed, if one excludes the “genre” topics, very few topics connect with more than two other topics. The nature-oriented topic 140 (landscape descriptions) is an interesting exception: it shows up in the first period co-occurring with topic 11 (traffic roads), topic 68 (Latin American Indians) and topic 69 (temples). Topic 140 is also of interest because it is the only nature-oriented topic – excluding “genre” topics and topic 163 (philosophy) – that is situated in the cultural cluster.

Identifying the specific Courier pages on which topic 140 (landscape descriptions) co-occurs with other topics in the network offers a means of charting in closer detail how the concept of nature intertwined with discussions of cultural expressions and other human activities. In the period 1948–1961, the topic of landscape descriptions connects to topic 11 (traffic roads) on pages about railroad-and highway-building. “Road teams are taming the wilderness”, reads one 1957 article, “stretching the frontiers of civilization, and bringing remote and hitherto untapped areas of the Earth's natural riches under development” [13]. Topic 140 connects also to topic 68 in articles about the way of life of indigenous communities (“Indians”) in Latin America. It co-occurs with topic 69 (temples) in articles like a 1955 discussion of St. Catherine’s Monastery in Egypt, which is surrounded by mountain peaks: “The savage beauty of this rugged, bare landscape is heightened by the mystery which surrounds the silent rocks” [14]. These examples show how, in the first period, landscape descriptions portrayed nature as wilderness, in contrast to “civilization” and “development”, or as the mystical, aestheticized setting of non-Western ways of life. The connection to nature’s “savage beauty” and “mystery” is part of what distinguishes these ways of life from modern urban life.

In the second period, 1962–1986, the distinctive words of landscape description appear in the context of natural sites in need of preservation, more or less without human presence, as we see in articles from this period in which topic 140 connects to the topics of water (49) and geology (63), for example in texts about waterways in World Heritage sites and about volcano parks [15]. In the third period, landscape language still links to preservation, but now in a manner that more readily links nature to culture: the pages on which topic 140 connects to topic 106 (national parks) come from articles about parks and nature reserves, in which we find references to the “long and extraordinary interplay between man and nature”, and “the aim of harmonizing the needs of people and nature” [16].

On the whole, however, nature-oriented topics appear in the three networks most often in the cluster of topics related to UNESCO and its programs, or to those having to do with development. Nature, in these parts of the networks, appears to be the subject of preservation or intervention for human welfare and development, as we see in figure 3, where “development” is the topic that links topic 49 (water), topic 36 (science), topic 106 (national parks) and topic 104 (climate change).
This form of topic network analysis is helpful, then, in identifying trends in the relationships among contexts in which a particular concept was discussed in the corpus. This mode of analysis can be useful for conceptual history without claiming that the topics themselves represent concepts or discourses. It is enough to accept that the computer-generated topics reflect linguistic patterns in the corpus, some of which are meaningful. We identify our target concept in a quite traditional way, by following its key words (the noun “nature” and the adjective “natural”) and using those words’ appearance in the computer-generated topics to visualize the relationships among the linguistic contexts in which the concept was invoked, as well as to measure how those contexts and relationships changed over time. This network approach also offers a means of identifying particular passages of interest for close reading, as we have seen. One drawback of the topic network approach is that many topics of interest, like for example topic 193 (human species), do not appear, either because they do not cross the topic weight threshold, or because they were apparently not discussed sufficiently often alongside other (sufficiently strong) topics. Our third approach to topic modeling is an effort to chart such topics, creating a novel means of engaging with the corpus.

5. Using topic modeling to identify thematic paths through the corpus

Here we use the topic model to identify those pages in the magazine in which a chosen topic is particularly strong in a particular time period. Doing this for several of the most intriguing topics produces a set of thematically sequenced pages across time, for manual reading and analysis. Here we explore one example, following topic 193 (human species) through the three time periods designated above. Beginning as it does with the words “human”, “man”, “nature”, “life”, “natural”, “species” and “environment”, this topic offers a compelling way to chart historically the uses of the concept of nature (expressed through the noun “nature” and the adjective “natural”) in the corpus. Moreover, the topic’s inclusion of the words “human” and “man” suggests an interest in the relationship with the realm of culture, broadly construed. Following one computer-generated topic will not of course reflect all uses of the concept. But that is the strength of the approach: rather than giving us a generic sample of places where the word appeared, this topic points us toward one specific linguistic context in which our target words were invoked.

The topic modeling algorithm we employ here enables us to identify the pages (in a given time period) on which a selected topic was strongest. (The relevant measure is of “topic weight”, a value that is normalized to the number of words on a given page.) We selected the first three pages (per period) for manual reading. What we found by reading articles in which topic 193 was strongest was a striking set of changes over time in the relationship between the concepts of nature and culture across our selected time periods.

In Courier’s early years (1948–1961), the articles in which topic 193 was especially strong emphasized a sharp distinction between nature and culture. These articles suggest that authors writing in Courier did this in part because that distinction was the core of an antiracist vision of humanity, inspired by contemporary trends in anthropology. A central point of post-1945 international antiracism was to argue that what accounted for differences among human groups, in terms of their ways of life or their apparent differences in social or intellectual aptitude, was not nature (biology, genetics), but culture. We see these arguments in the earliest articles highlighted by following topic 193, like a 1953 article by the Rutgers University anthropologist Ashley Montagu on the idea of “human nature”. Here Montagu argues that man is “the most plastic, the most malleable, the most educable, of all living creatures” because our real differences are not a matter of nature, just culture. “Science”, Montagu writes, “knows of no natural drive in man to make knives and forks or to speak Italian; Australian aborigines neither use knives or forks nor do they speak Italian, not because they couldn't do so, but because they happen to be born into a cultural environment from which such instruments are absent and where their own language alone is spoken” [17]. Similar invocations of a strong dichotomy between nature and culture can be found early in the following period (1962–1986). There, topic 193 leads us for example to a 1965 article by the Belgian researcher Jean Hiernaux on “cultural evolution” – namely, the progress that humankind has made and makes by accumulating and passing down knowledge and
skills, not through genetic change. Here, too, we find the concepts of nature and culture discussed in close relationship to one another, and again, more or less as opposites [18].

In the second period (1962–1986), we see signs of a different nature discourse, as topic 193 leads us to articles arguing that nature and culture are (or should be) in a delicate balance – not opposites, but integrated. Topic 193’s strongest appearances in period two include a 1969 article by René Dubos, in which the French-born American microbiologist and prominent public intellectual calls for a new mix of human agency (culture) and the physical, non-human environment (nature), ideally resulting in what he refers to as “civilized nature”. Motivating this call for a balanced integration of nature and culture is Dubos’s doubt that the two spheres can really be understood in isolation from one another. We see this where he comments explicitly on the weakness of the prevailing understanding of “nature”: “The ill-defined meaning of the word Nature compounds the difficulty of formulating a scientific basis for the philosophy of conservation. If we mean by nature the environment as it would exist in the absence of man, then very little of it survives” [19]. Nature wholly distinct from culture is conceivable to Dubos, but so rare as to be almost irrelevant.

In the third period, beginning in 1987, topic 193 identifies discussions in which new ways of thinking about nature are pushing into areas that were once purely the domain of culture. A 1993 interview with the French philosopher Luc Ferry discusses the notion that nature might have rights, and the question “to what extent does ecology question the basis of modern civilization” [20]. Articles like this one gave voice to doubts about the possibility or desirability of what Dubos had called “civilized nature”. Now there is concern that nature must be understood to have systems of its own (embraced by the newly popular category of ecology) as well as values or even rights of its own, apart from any human cultural viewpoint.

But the most recent articles identified by following topic 193 show still another turn. This same topic leads us to a 2018 interview in which the historian Dipesh Chakrabarty introduced Courier’s readers to the concept of the Anthropocene, the name recently suggested for our current geological age, characterized by humankind’s impact on planetary systems, including above all the climate. Here Chakrabarty argues that “humans are a geological force”, such that nature and culture cannot be separated even at the level of the planet [21].

This brief tour of Courier through topic 193 shows us that Chakrabarty’s point was not so new: Courier authors had suggested the interconnection of culture and nature back in the 1960s, albeit in less radical terms. We see, moreover, that what was at stake in arguing for or against a sharp nature-culture distinction changed a great deal over the last half-century. In the 1950s, progressive voices at UNESCO insisted on a sharp nature-culture distinction as part of the fight against racism and for a new world order of human equality. By the early twenty-first century, it seemed like a matter of great urgency that we abandon a sharp nature-culture distinction, so that we could understand and act on the existential threat to humanity posed by climate breakdown.

This was just one of multiple such explorations of the corpus that this approach permits. It is an exploration that is, ultimately, old-fashioned in its focus on reading and contextualizing. But it was driven by a combination of word-searching with algorithmically generated topics, and would not have worked otherwise.

6. Using topic modeling to identify thematic paths through the corpus

The research documented in this article leads us to two conclusions, one methodological and one empirical. Regarding methodology: we have explored three modes of using topic modeling for conceptual history. The selected approaches seem to work better, or at any rate differently, at different scales. Each scale offers a different perspective on the corpus, from the most distant (topic networks) to the closest (reading pages identified by following topics of relevance to a chosen concept). We find, too, that the approaches complement one another. Observations we make by reading passages in which a topic is particularly strong can be put into perspective by the quantitative data that comes from having modeled the entire corpus. We know, for example, that topic 193, while interesting, was not terribly strong in the corpus as a whole (in terms of topic weight it ranks at place 83 out of 200). On the other hand, our findings from close reading articles signaled by this topic offer perspective on the findings
we made through the more “distant” approaches used earlier in the paper, in the sense that we find that a nature topic that seemed to be focused on nature as the object of scientific research was, in fact, a highly significant context for discussions of the nature-culture relationship.

As for an empirical conclusion: the concepts of nature and culture engaged in a complex dance in *Courier*, now closer, now further apart, and the journal seems to offer rich resources in which to follow it. Judging from the results obtained by following topic 193, the dance seems to have been led by European thinkers, with the important exception of the Indian-born historian Dipesh Chakrabarty. Does this mean that the global character of the magazine did not amount to much, as far as this theme is concerned? That will be an interesting question to explore. We are encouraged that topic modeling will offer one stimulating means of doing so.

**Acknowledgments**

Research for this paper was supported by a grant from the Swedish Research Council for the project “International Ideas at UNESCO: Digital Approaches to Global Conceptual History” (VR 2019-03278). Thanks to the UNESCO Archives, and especially archivist Eng Sengsavang, for facilitating access to document. Many thanks also to Roger Mähler and Andreas Marklund of Humlab (Umeå University) for their invaluable research support.

**References**


