Journal of Field Archaeology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/yjfa20

Excavating Archaeological Texts: Applying Digital Humanities to the Study of Archaeological Thought and Banal Nationalism

Gertjan Plets, Pim Huijnen & David van Oeveren

To cite this article: Gertjan Plets, Pim Huijnen & David van Oeveren (2021) Excavating Archaeological Texts: Applying Digital Humanities to the Study of Archaeological Thought and Banal Nationalism, Journal of Field Archaeology, 46:5, 289-302, DOI: 10.1080/00934690.2021.1899889

To link to this article: https://doi.org/10.1080/00934690.2021.1899889

Published online: 22 Mar 2021.

JOURNAL OF FIELD ARCHAEOLOGY
2021, VOL. 46, NO. 5, 289–302
https://doi.org/10.1080/00934690.2021.1899889

Excavating Archaeological Texts: Applying Digital Humanities to the Study of Archaeological Thought and Banal Nationalism

Gertjan Plets[a], Pim Huijnen[a], and David van Oeveren[b]

[a]Utrecht University, Utrecht, The Netherlands; [b]Maastricht University, Maastricht, The Netherlands

ABSTRACT

To date, the evolution of archaeological knowledge production and theory has been discussed and analyzed using qualitative methods by reading vast amounts of archaeological texts in search of specific discourses or framings of the past. In this paper, we present text mining methodologies from digital humanities that can be applied to large corpora of archaeological texts to trace and evaluate changing knowledge practices. Such a big data approach is imperative. Due to the rapid increase of archaeological publications, qualitative research into the intellectual history of archaeology has become complicated and highly selective. The big data methods presented in this study were tested on a large corpus (4,811 texts totaling over 51 million words) of different types of archaeological texts from the Dutch-speaking part of Belgium. The different text mining tools were successful in identifying theoretical trends. Our tools were also successful in charting the decrease in quality due to changed organizational circumstances (developer-led archaeology). Furthermore, we could also map changing banal nationalist framings of the past.

KEYWORDS

history of archaeology; archaeological theory; text mining; Belgium; nationalism

Introduction

Archaeological literature has been a principal source to reconstruct and explore the cultural and political dimensions of archaeological knowledge production. Since we spend as much time writing and reading as doing fieldwork, texts are not only central in the production and circulation of archaeological knowledge (Hodder 1989; Lucas 2018), our socioculturally constructed frames of thought are also encoded in them. Using approaches from the history of science, archaeological texts have been successfully deconstructed in seminal intellectual histories of archaeology to discern dominant ideas and evaluate the context of the archaeologist (Diaz-Andreu 2008; Johnson 2019; Trigger 2006).

Qualitative methods have stood central in these efforts. This means that textual sources are subjected to close reading to unravel those structuring discourses encoded in archaeological monographs, reports, and articles. Although this approach is detailed and effective, key interpretations are often based only on a limited number of texts and not all types of literature produced in the archaeological social arena. In these studies, monographs and important journal series especially occupy a central role, ultimately favoring archaeologies produced by academic archaeologists and especially those protagonists structuring key debates. However, since the mid-20th century, the archaeological field has diversified and become populated by different types of archaeologists that actively influence how archaeological knowledge is produced and how specific regimes of truth are normalized. These archaeologists' perspectives have found their way into the numerous new journals and book series that have been launched in the past half-century. This proliferation in archaeological text further accelerated by the end of the 1980s, with digital text formats and developer-led archaeology.

This abundance in players and texts complicates the systematic analysis of all archaeological literature. Consequently, most studies to date limit themselves to either traditional academic textual products such as monographs and established journals (Kristiansen 2012) or study only a small—often randomly selected—sample (Börjesson 2015). The continuing focus on academic monographs is especially problematic. All different formats, ranging from articles to reports, define archaeologists' and ordinary people's understanding of the past (Lucas 2018; Seymour 2010). Insights from the history of science teach us that, within studies of knowledge production, it is important to understand the whole thought collective, since all actors interact and share insights. In short, it is important to analyze the entire textual corpus.

In this paper, we present a new approach to the study of archaeological thought by "excavating" archaeological texts with computational methods. Using digital text mining, we present a digital workflow for the diachronic study of large text repositories. The advantages of digital techniques are, firstly, the ability to analyze larger quantities of text over longer periods. Secondly, they enable scholars to find patterns and trends latently present that are sometimes impossible to discern on the level of close reading (Morretti 2013). In this study, basic text mining tools are presented to trace 1) theoretical trends, 2) the impact of organizational circumstances on knowledge production, and 3) nationalist framings of the past. These three components are traditionally seen as structuring the archaeologist's context (Hodder 1991; Trigger 2006).

The application of text mining methods to archaeological corpora is not new. Projects have used digital text analysis to index archaeological information and grey literature

CONTACT Gertjan Plets g.f.j.plets@uu.nl Department of History and Art History, Utrecht University, Drift 6, Utrecht, The Netherlands.

(Brandsen et al. 2019; Richards, Tudhope, and Vlachidis 2015). In these projects, digital text mining positions itself as a key tool to extract archaeological data out of the plethora of texts associated with contemporary archaeological practice. However, in this paper we contend that text mining should not only become a part of the archaeologist's toolkit to paint a better picture of the past, it should also become more widely used to study the history and evolution of archaeological knowledge production. In the field of history, especially, the study of knowledge production has benefited enormously from computational text tools. The mass digitization of source material and the introduction of new tools has enabled historians to analyze large amounts of unstructured texts. This allows historians to trace conceptual change over time, map the circulation of ideas (Hall, Jurafsky, and Manning 2008; Tangherlini 2013), and extract the changing interpretations of ambiguous concepts (e.g. de Bolla 2013; Guldi 2019).

This paper positions itself as a proof of concept for the application of such digital history and text mining methods to the study of archaeological thought. Both standard and more refined tools are applied to a vast corpus of diverse archaeological texts to trace basic—albeit important—evolutions in archaeological practice, theory, and methodology. Consequently, this article should be seen as a first step towards the more widespread application of digital humanities methods to the study of the history of archaeology. We acknowledge that much more complex tools can be used to study very specific developments. Therefore, we hope that the baseline developed by this paper will encourage such studies and revitalize the study of archaeological thought. Finally, we wish to emphasize that the methods proposed should not replace older, established, qualitative approaches, but should rather be used in combination with close reading and hermeneutics.

This paper is built around the workflow of a digital history project. In the sections below, we will describe how a dataset needs to be built, which tools can be employed, and how the results should and can be interpreted. In the last section, we bring the insights developed through our big data approach in conversation with other, more traditional, studies of archaeological thought and politics. The presented digital history approach is based on the exploration of all archaeological literature on the territory of Belgium produced during the period 1945–2017. Only literature in Dutch was selected, since multilingual analysis would have gone beyond the scope of the article. A total of 4,811 documents were included, totaling over 51 million words. All text types were assessed, ranging from monographs and journal articles to excavation reports.

Belgium was strategically selected as a case study for several reasons. First, the Dutch-language literature from Belgium was expansive enough to provide a sufficient dataset for diachronic analysis, while being small enough to ensure that all archaeological texts could be included. Second, Belgium's political and institutional landscape has drastically evolved since 1945. From the 1980s onwards, the country has evolved into a federal state in which the regions (Flanders, Wallonia, and Brussels) govern over heritage (Van Looveren 2014). In these regions—mainly Flanders—strong nationalist movements define cultural fields of practice. In addition to nationalism, administrative reorganizations in connection to cultural resource management have drastically impacted how archaeology is conducted within academia, the private sector, and public institutions. As a result, through our dataset we can trace how changing administrative regimes define heritage discourses (cf. Geismar 2015; Plets 2016).

Building the Dataset

Collection

Since evolution in archaeological knowledge and thought stands central in this study, only literary formats in which archaeologists interpret and analyze archaeological findings were included. For the period 1945–2017, all Dutch language journal articles, book series, monographs, Ph.D. theses, and excavation/survey reports were collected. Reports produced within the context of developer-led archaeology were also included, since they have become a cornerstone of the archaeological field and are an important source used in the development of archaeological knowledge (Aitchison 2010; Evans 2015). Short discovery notes were excluded because they lack analytical depth. Since we wanted to study changes and discrepancies in archaeological knowledge, attributes were registered for each text: date of publication, nature of publication (monograph, journal article, CRM reports, or conference proceedings), and type of player (national government, regional government, academic, and commercial).

Just as most European heritage agencies, the Flemish Heritage Agency has created a digital archive (https://oar.onroerenderfgoed.be) holding all excavation/survey reports and journal/monograph series published by the agency. Since mostly reports and monographs from 1990 onwards were available on the portal, older materials and those works produced by other players needed to be digitized. Due to the vastness of the textual archives, all papers were photographed with a standard DSLR camera. The ABBYY FineReader 15 platform was selected for optical character recognition (OCR) in these files.

Processing

From the collected data, all English or French text (e.g. summaries) was automatically deselected. Texts with significant overlap (e.g. reprints) were also removed. As a result, about 5% of the dataset was dismissed, resulting in 4,696 titles. These texts were preprocessed for further analysis based on a bag-of-words approach. This means that all textual data was stripped of interpunction, numbers, and HTML markup; the entire corpus was set to lowercase; words shorter than three letters were removed; and, all standard stop words were removed from the corpus. This was done to improve the processing speed and mitigate the large number of brief non-words that OCR-ing of digitized material usually produces.

The cleaned corpus consisted of more than 22 million words. Since measuring changes over time was a key ambition of this project, a lot of effort was put into collecting material from earlier periods. This was not straightforward, since archaeology in Belgium was not fully established in the mid-20th century, and many works were still published in French. Although the corpus becomes richer from the mid-1970s onwards (Figure 1), earlier texts can still be

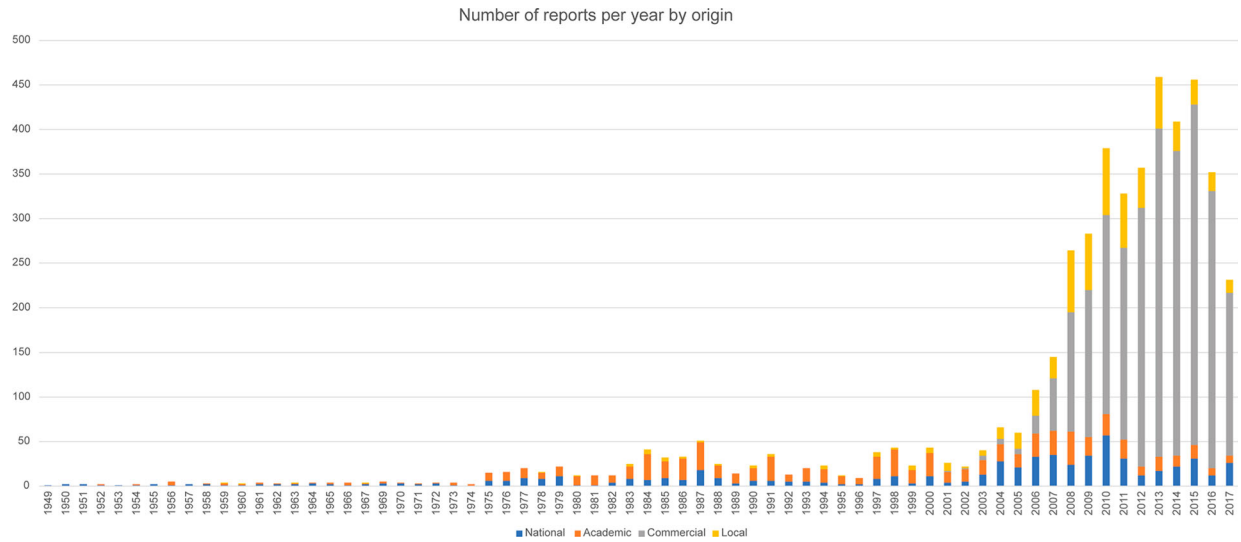

Figure 1. Corpus statistics indicating the number of texts per year per player.

used to explore basic trends and evolutions. Since we want to encourage experimentation with our data, replication of our results, and comparative analysis with other (national) contexts, our data set was made available (open access) for reuse (Plets, Huijnen, and van Oeveren 2021).

Analysis

Mapping changing theoretical schools: concept mining

The questions we ask, and ways we order data to build interpretations of the past, are strongly influenced by the theoretical frameworks we deliberately or unintentionally deploy (Johnson 2019, 2). Many archaeologists might claim that they "don't do theory," but the reality is that we always employ multiple analytical lenses and preconceptions through which we look at the data produced from fieldwork, or on an even broader level, how we select data and sites for excavation. Therefore, discussions about the history of archaeological thought have over the past decades especially drawn attention to the specific archaeological theoretical traditions (Hodder 1991; Trigger 2006) or broader philosophical ontologies (Thomas 2004) that have or continue to texture archaeological reasoning. A rich variety in theoretical sub-schools have developed in the last half-century. Nevertheless, most standard works on archaeological theory (Johnson 2019; Trigger 2006) still connect these numerous individual schools to a select number of archaeological traditions that are connected to broader philosophical and intellectual movements, such as cultural evolutionism, positivism, and postmodernism.

We are immensely interested in tracing these sub-traditions and individual schools in the Belgian corpus and believe it is essential to develop a fine-grained understanding of the history of Belgian archaeological thought. However, since the ambition of this paper is more modest and seeks to study and evaluate the application of methods from digital history to the history of archaeology, we will similarly only focus on those established archaeological traditions: culture-historical archaeology, processual archaeology, and post-processual archaeology. In our study, we will evaluate which core theoretical frameworks define and structure archaeological propositions in Belgium and how these evolve over time. Drawing on discussions in Science and Technology Studies (STS) and the history of science, we underline that these traditions should not be seen as mutually exclusive paradigms that supplant each other. Rather, core ideas and assumptions within these schools are often employed at the same time to varying degrees across Europe (Hodder 1991, 11–12). As such, we do not intend to create a periodization of different theoretical paradigms, but rather aim to trace which intellectual frameworks circulate in the actor-network of relations in Belgium and which traditions have been more central than others at various stages in Belgian post-war history.

Digital humanities scholars have been experimenting with machine learning approaches to study continuities and changes in disciplinary schools of thought, in particular using topic modeling (see Tahmasebi, Borin, and Jatowt 2018 for an overview). Topic modeling makes visible semantic co-occurrences between words that are often used in conjunction and, thus, together form latent clusters. This technique has been used to study the thematic evolution of, for example, scholarly disciplines (Hall, Jurafsky, and Manning 2008) and political discourse (Guldi 2019).

The MALLET package was applied to our corpus (see acknowledgements for settings and links to used programs). Despite the large variety of settings used (like the number of topics and the hyperparameter optimization, as well as different segmentations of the input data), this did not result in significant changes over time. Most thematic clusters were built around descriptive language that remained similar over seventy years. Clearly, an empirical and descriptive discourse defines how the archaeological record is reported throughout the different years of the corpus. One could argue that interpretations on a meta-conceptual level are not central in the writing culture, and therefore Dutch-language Belgian archaeology is atheoretical. However, as noted by Johnston, empiricism and the dominant belief that "facts speak for themselves" (Johnson 2011, 765) should itself be seen as a formal theoretical position. Others (Gramsch and Sommer 2011; Hodder 1991) argue that empiricism is especially widespread in continental Europe and is tangentially connected to the still popular culture-

historical tradition where collection, cartographic plotting, and ordering of data stands central (Trigger 2006).

The absence of specific theoretic clusters in our topic models, besides the empirical prose, does not mean there are no influences from key schools of thought. Rather, it points to the limited direct use of these theories, meaning that there are few explicit theoretical discussions in Belgian texts. Implicit references to dominant ideas can still be discerned from large corpora, even when topic clusters are absent. Archaeologists, like any other scientists, heavily draw on concepts when assembling knowledge about the past. Concepts are individual words that operate as signifiers in which intellectual and interpretative frameworks are encoded. Consequently, theoretical traditions can also be identified by tracing the use of specialist vocabularies structured around timely and fashionable concepts. In this study, therefore, the evolution in the application of dominant ideas associated with the three aforementioned schools was studied through a concept historical approach. The evolution in the use of key concepts was quantitatively mapped through developing a thesaurus of theory-specific wordings and jargon. It is a proven technique to create a thesaurus of key terms based on relevant input data and domain knowledge. Digital humanities scholars have used this technique, for example, to trace neoliberal discourse over time in German public opinion (Wiedemann, Lemke, and Niekler 2013).

Within archaeology, there is an extensive Anglophone literature describing the different archaeological traditions and associated concepts. A Dutch language thesaurus of theory-specific concepts was, unfortunately, not available for this study. Consequently, a list of key concepts and phrasings (Table 1) was made by subjecting a selection of texts to close reading. Wordings and jargon specific to major archaeological interpretative frameworks were selected. Importantly, concepts were selected that are otherwise not widely used in the Dutch language. Works by key Dutch-speaking Belgian archaeologists (Crombe 1996; De Clercq 2009; de Laet and Glasbergen 1959; de Laet 1974; Thoen 1987; Tys 2004) known for—albeit implicitly—mobilizing key archaeological frames of thought were selected for qualitative hermeneutic analysis. These works were not included in the corpus analyzed for changing theoretical propositions, since they would be overrepresented and heavily influence the final results. The used works are either dissertations, monographs, or expansive articles by leading professors in archaeology who influence(d) the field.

Using Voyant Tools, the frequency of these words was measured for all years. Subsequently, for each school, the ten most used concepts (and their allied words and synonyms) were selected for comparison and diachronic analysis. These ten most used concepts, expressed in frequency relative to the total number of tokens per year, were plotted on a graph to find patterns and trends within the relative popularity of the schools of thought during the period of study.

Table 1. Thesaurus of the ten concepts per theoretical school for comparison (and their synonyms or allied concepts). Words with * are root words: for these words we also counted the words to which a prefix or suffix was attached.

Culture-Historical Archaeology	Processual Archaeology	Post-Processual Archaeology
Beschaving*	Klimaat, Klimatologisch, klimaatsverandering, Afkoel*, Vernatting	Agency
Diffusie	Bodemkund*, Pedologi*	Etniciteit
Franken, Frankisch	Petrografi*	Gender
Gallo-Belg*, Belgae, Eburonen, Eburoon, Nervi*	Biodiversiteit, Biomassa, Vegetati*	Groepsidentiteit, Identiteit
Germaans, Germanen	Ecologi*	Seksualiteit
Kelt*	Fysisch-geografisch*	Habitus
Kulturen, Culturen, Kultuurprovincie,	Landschap, Landschappelijk	Sociale structu*
Urnenveldenkultuur, Urneveldekultuur Volk, Volkeren	Overexploitatie	Macht, machtsrelatie*
Migratie, Volksbeweging, Volksverhuis, Volksverhuizing	Statisti*	Materialiteit

Studying the impact of organizational circumstances: institutional organization and research quality

For most of their history, museums, universities, and national heritage agencies presented themselves as scholarly centers of archaeological knowledge production. With organizational circumstances comparable to other disciplines within the humanities, it is logical that discussions over theory and methods presented themselves as the main heuristics to explore archaeology's intellectual history (Thomas 2004; Trigger 2006). However, this changed with the advent of commercial "rescue" archaeology. In Europe specifically, the Valletta convention dramatically reshaped the archaeological landscape.

The archaeological community agrees that this organizational change has reshaped the entire discipline's research culture and understanding of the raison d'être of archaeology (Carver 2007; Kristiansen 2009). Although some point to the new questions that can be asked with the produced "big data" (Vander Linden and Webley 2013), most debate centers around the changing research standards developer-led archaeology has enacted both in cultural resource management and the field more broadly (De Clercq et al. 2013; Demoule 2012). This discussion has inescapably been normative in nature, meaning that discussions on changing research standards have centered around questions of quality. In discussions on the relationship between research standards and organizational circumstances, "socialist/regulated" regimes where the public sector plays a central role have been weighed against "capitalist/market-regulated," preventive archaeology where market-driven processes govern and define the rules of the game (Carver 2007; Kristiansen 2009; Willems and van den Dries 2007).

In search of an organizational circumstance in which developer-led archaeology results in both data *and* information profit for both society and academia, three dimensions to quality have often been foregrounded: 1) research methodology and design—are appropriate registration methods used, and is the research method tailored to local circumstances or are standardized strategies used (De Clercq et al. 2013; Roth 2010); 2) theoretical framework—archaeology is always more than the basic recovery of material evidence; it needs a conceptual and comparative component where the data is contextualized and interpreted (Champion 1991); and, 3) scholarly interaction—engagement with historiography is not only crucial for the development of a research question, but the archaeological past should always be comparatively studied through engagement with other literature (Kristiansen 2001).

It is true that scientific quality is much more holistic and cannot be distilled to a select number of characteristics. Still,

the above principles that focus on historiography, method, and theory deal with the bedrock of academic research. However, measuring these standards remains incredibly difficult, and the debate lacks clear metrics. A few researchers have used biometric analysis to study the degree of scholarly interaction in archaeological literature. Manual (Kristiansen 2001, 2012) and recently also semi-automated methods (Börjesson 2015) have been developed to assess citation patterns. These contributions point at decreasing citation practices as preventive archaeology becomes dominant. Since we also wanted to explore the other two components of research quality, we selected additional proxies to evaluate the methodological and theoretical soundness of a project.

Tailored methods or boiler-templating?

Scholars signaled that "elaborateness" (i.e. length of the report and research design section) can be an indication for well-thought-out, apt methodological frameworks (Evans 2015). At the same time, Roth (2010) has lamented the rise of long, standardized research designs and method sections in archaeology and tied it to heavily standardized practices on the ground. A phenomenon that many of us have witnessed is, according to Roth, that excavation reports have become nothing more than "boiler template documents" (Roth 2010, 340) with limited attention to local needs and historiographical discussions. Such a practice would mean that archaeological excavation and reporting has become heavily standardized and has little attention for academic debates concerning specific periods or regions, as well as a less contextual methodology and excavation strategy tailored to specific geographic settings. This study uses two metrics from digital linguistics to investigate the extent to which boiler templates have become common practice: 1) cosine distance and 2) text reuse.

The cosine distance technique is based on the vectorization of texts. In our case, this means that the frequency of every unique word in our texts-per-year provides the coordinates in a multi-dimensional space. In this way, each year can be represented by a vector that starts at [0,0] and crosses this coordinate. This technique measures the cosine of the angle between two vectors representing two different years, where 1 equals absolute similarity and -1 absolute difference. Cosine distance, which we have calculated with the popular Python Scikit-Learn package, gives us an indication of the degree of overlap in wordings between years (Aggarwal 2015, 75–76).

An additional metric was selected to quantify the text reuse between subsequent years by using the reuse detection tool Text matcher. This allowed us to hone in on the absolute similarities between archaeological reports. The set threshold and the Ngram value we matched against for plagiarism enabled us to detect text reuses in which just a few words were changed or had faulty OCR readings. Only the reuse of blocks of 220 or more characters were studied, since we wished to exclude the repetition of bibliographic entries.

Analytical depth: lexical complexity

As a general rule, texts in which data is comparatively studied and theoretically explored on a meta-conceptual level are generally characterized by a more complex lexicon than those in empirical descriptive prose. In computational linguistics, a suite of tools has been developed to measure and evaluate lexical sophistication and writing quality (Kim, Crossley, and Kyle 2018). Since it is the ambition of this paper to introduce basic methodologies from digital humanities, this study limits itself to two basic proxies: 1) lexical density and 2) the use of specialist language.

Linguists (Halliday 1989; Johansson 2009) argue that lexically dense texts are rich in different information-carrying words. More descriptive texts have a lower lexical density. We evaluated this lexical density by counting the proportion of unique words (types) against the total number of words (tokens) in all texts per year, differentiated by the type of publisher. This well-known type-token ratio (TTR) has the weakness that it does not account for differences in text length. This is no trivial flaw. After all, the number of unique words will not increase proportionally with overall text length.

To counter this, we have used the standardized type-token ratio (STTR), which subdivides the texts-per-year into smaller chunks (in our case, of 1,000 words) over which the TTR is computed. The STTR is, then, the mean of all TTRs of one year. From the STTR, the TTR of the chunks with less than 100 types are excluded because these segments create a negative bias in the STTRs. These are particularly prevalent in our corpus, since archaeological texts are rich in tables, lists, or captions rather than running texts.

Lastly, another proxy for analytical depth is the presence of sophisticated words that are "less concrete, less imaginable and less familiar" (Kim, Crossley, and Kyle 2018, 121). Academic jargon and theoretical concepts fall under this category. Since we already developed an inventory of theoretical concepts (see above), the frequency of theoretical jargon per year was used as a proxy for analytical depth.

Evaluating the impact of the political field: tracing latent nationalism

Although the field of politics is complex and archaeologists play an active role in normalizing a host of political structures and agendas, national identity in particular remains a key political construct that archaeologists both actively shape and are shaped by. Many archaeologists unconsciously assemble modern nations through often identity-laden findings and artifacts that become appropriated by nation-builders. At the same time, seminal political analyses (Meskell 1998; Niklasson 2013) of archaeology have taught us that archaeologists are, at the same time, also being shaped by modern ideas of the nation and include these frames in their research questions or orderings of the past. Those examples we know best from literature are the more extreme cases from the 19th–early 20th century, where archaeologists explicitly tied artifacts to national origin stories or explored the past through the lens of 19th century identity constructions (Diaz-Andreu 2008; Dietler 1994).

These more outright nationalist archaeologies have ensured that, over the past decades, archaeologists have become aware of the politics of heritage. Consequently, explicit reifications of the nations and the reproduction of ethno-historical myths have become less common. However, this does not mean that the nation is not latently present in most archaeological discourse, and everyday archaeologists indirectly reproduce and assemble contemporary national identities and provide them with cultural time-depth.

Discussions within social science remind us of the widespread nature and strong impact of so-called methodological

nationalism. Methodological nationalism can be best described as the "assumption that the nation/state/society is the natural social and political form of the modern world" (Wimmer and Glick Schiller 2002, 301), meaning that the nation-state is too often used as the dominant vector of analysis in describing and analyzing processes in past and present. Building on Billig (1995), Wimmer and Glick Schiller argue that this is not without repercussions, since by routinely using the nation-state in an almost banal way in scholarly discourse, present-day national imaginations become naturalized.

In archaeology, discussions about methodological nationalism are limited (Niklasson 2013; Plets 2016). However, in the allied field of history, it is more common to explore latent nationalist framings and to study how contemporary nationalist constructions become unintentionally interwoven in historical interpretations (Brown and Davis-Brown 1998; Torsti 2004). Discussions within the subfield of transnational history in particular have argued that methodological nationalism in studies of the human past manifests itself in the demarcation of study regions or in the selection of data and literature used for comparative analysis. As such, present-day borders influence how we study the past. This is not without broader societal effects. Methodological nationalism normalizes present-day administrative regions as meaningful historical analytical containers, providing contemporary nations with deep cultural anchors that further naturalize them as primordial historic homelands.

Since methodological nationalism manifests itself in an implied way, equally indirect proxies were selected to quantitatively approach methodological nationalism. During close reading of various archaeological texts and registering the titles of the thousands of articles and reports studied, the authors noticed that contemporary signifiers were often used in conjunction with archaeological periods. These were often epochs for which present-day administrative signifiers had little or no meaning. For example, earlier texts would colloquially talk about the "Belgian" Bronze Age. More recent works would, in turn, describe the relevance of their excavation for better understanding "Flanders" in, for example, the Roman period. Both statements might not be blatant nationalist instantiations firmly anchoring the nation into the past; still, present-day administrative borders are used in a meaningful way to describe and discursively assemble the past. Such expressions might not directly essentialize the nation, but, as everyday speech acts, they could point at changing ethnocultural frames of reference used by people in making sense of the world around them (cf. Billig 1995).

In an effort to trace such indirect banal nationalist discourses, we mapped the frequency of specific geographical signifiers used strictly within descriptions and interpretations of the past. Only concepts that are interwoven with identarian values (Belgium, Europe, and Flanders) were selected. Since this study only wished to explore spatializations of the archaeological past, all concordances that could be tied to administrative institutions (like Agentschap voor Geografische Informatie Vlaanderen), legislation, information infrastructures (e.g. cartographic portals like geovlaanderen.agiv.be or gisvlaanderen.be), websites (like belgica.kbr.be), titles of journal and monograph series, and names of people (e.g. Duke of Flanders) were omitted from this calculation. Next, in Voyant Tools, the frequency of signifiers relating to Flanders ("vlaams*", "vlaan*"), Belgium ("belgi*"), and Europe ("europ*") were counted. We stress that we employ this proxy as an indirect indication of methodological/banal nationalism. As in any study of nationalism, further qualitative textual research is necessary, ideally backed up with ethnographic research amongst the study group.

Results

Theoretical schools

Word frequency analysis of theoretical concepts (Figure 2) shows that culture-historical phrasings are latently present throughout the entire study period. While concepts from other schools are largely absent up to 1980, culture-history seems to be of structuring theoretical influence. Together with the empirical nature of most texts (see above), the presence of key concepts related to cultural evolutionism confirms the importance of this tradition in Belgium. There are two significant peaks in the use of culture-historical signifiers, indicating also more manifest engagement with these theories: during 1955–1967 and 1980–2004. Explanations for these peaks can be connected to theoretical developments in Belgian academia and changing organizational circumstances (see below).

The first peak might be tied to the publication of two syntheses in the 1950s on the (pre)history of Belgium by the influential archaeologists de Laet and Glasbergen (1959) and Marien (1952). Trends in the corpus could indicate that peers engaged with these monographs—perhaps especially the 1959 book by de Laet who was a professor in "national archaeology" at the main Flemish university in Ghent. Both works are heavily prescribed culture-historical explanatory models (De Mulder and Bourgeois 2017). A much more significant growth can be discerned starting in the late 1970s that peaks in the 1990s–early 2000s; this peak hints that archaeologists more explicitly engaged with culture-historical notions of (pre)history during this period. A few archaeologists we interviewed indicated that especially since the mid-1980s, they became more interested in using theory. Many stated that due to increased possibilities for international mobility, they would more frequently attend theoretical conferences in the Netherlands (e.g. Archaeological Dialogues conferences), forcing them to include theoretical concepts and ideas in their presentations. It is plausible that this interaction with "theoretical archaeology" encouraged the application of more theoretical lenses when analyzing the archaeological record. The theoretical models employed were those that were already widespread in Belgium. Processual thought might have already been at the forefront in the Anglophone world, though culture-historical thought was still the defining theoretical approach in continental Europe (Gramsch and Sommer 2011).

Today, in the Dutch-language literature produced in Belgium, culture-historical concepts are still used frequently. Their frequency, however, decreased sharply in the mid-2000s. However, its continuing presence indicates a more implicit and latent engagement with this tradition. Also in other national settings in continental Europe, culture-historical frameworks continue to structure archaeological discourse (Johnson 2011), despite a clear shift away from the broader culture-historical project of tracing ethnic histories and movements of people. Rather, in those contexts, which

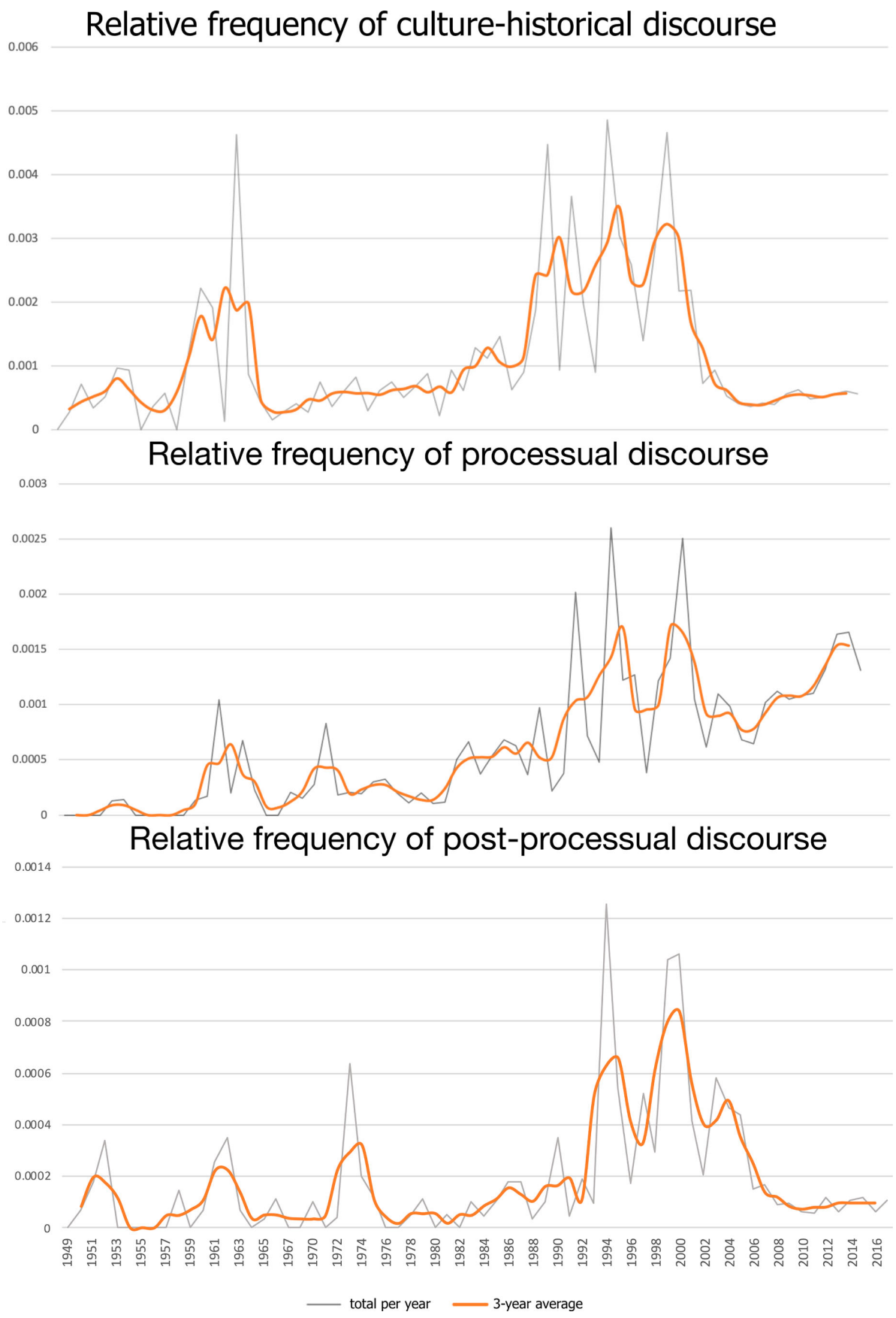

Figure 2. Graphs indicating the relative use per year of vocabulary tied to different archaeological schools.

is also the case in Belgium, culture-historical archaeology has a strong residual presence, still structuring how archaeological data is interpreted, albeit less explicitly.

Concepts tangentially connected to processualism appear in the mid-1980s and strongly increase in the 1990s. The appearance of concepts almost never used before indicates new ideas and frameworks are now used to describe archaeological data. The relative increase of these concepts indicates a manifest and more explicit engagement with these ideas, especially since the late 1980s. The initial rise in the

1980s might be tied to the fact that at that time processual theory was at its height in Anglophone archaeologies and structured international literature. Many interviewees indicated that during this period, literature from the UK and US was more commonly consulted, and works by scholars such as Colin Renfrew and Lewis Binford were especially popular. This sharp rise can also be explained by the widespread adaptation of methods and information infrastructures (e.g. GIS) from the hard sciences, encouraging a more positivist engagement with the past. Although the relative use of processual vocabulary decreases in the mid-2000s compared to other traditions, it continues to be employed on a more fundamental level and is now the main theoretical school. This might indicate that today, against the residual presence of culture-historical ideas, frameworks connected to processualism are relatively important.

Finally, post-processual vocabulary is seldomly used only during the second half of the 1990s–early 2000s. This overlaps with the international growth in seminal literature in this period. Its use abruptly stopped in 2004–2005 (except in the dissertation of De Clercq [2009] used for reference).

Lexical complexity and research quality

Exploring elaborateness and boiler templating

Evans (2015) hinted that over the past years, the quality of archaeological reports in the UK had increased because research strategy sections had become more elaborate and detailed. A similar trend can be discerned in our dataset, where reports became much longer than the years before (Figure 3). When looking at the individual reports, methods sections have indeed become longer, and choices in the field are elaborately described. However, one can have more words, but if these words are largely the same and descriptions about the archaeological record follow a very similar line of argumentation, the scholarly labor put into "ordering" data was marginal.

Since we can only compare the cosine similarity of one year with other years, we calculated the similarity, per player, between each year and the years before it, and the similarity between texts of 2017 and all preceding years. Since texts produced before 2003–2004 are very dissimilar, indicating that during this period (self) plagiarism did not occur, only texts produced in the last 15 years were included in the graphs (Figure 4). The outcome of the analysis clearly shows that with the rise of developer-led archaeology (see below), there is a very strong similarity in the words used to describe archaeological findings and develop interpretations. The graphs show that texts reproduced by commercial actors are especially very similar and share a lot of words between each other, indicating practices of boiler templating and (self) plagiarism. Other players' texts are less similar. However, between 2007 and 2011, the other actors' texts score higher in cosine similarity compared to other years. This is the period when public institutions and universities engaged in developer-led archaeology. Up to 2008–2009, most developer-led excavations were conducted by public institutions. After this period, commercial players started to dominate the field, and all other players only rarely engaged in this practice. When the latter players become less active, the cosine similarity also drops.

Text Matcher provided insight into how many times blocks of text of 220 characters (in which a few words were changed) were recycled from previous years (Figure 5). Similarly, up to 2003–2004, hardly any text reuse happened. From 2003–2004 onwards, more and more blocks of text from previous years were reused, and, especially in 2011, this practice became widespread and further increased. The produced data shows that copy-pasting with minor changes has become very widespread in Flemish archaeological texts. When public institutions were engaged in market-driven archaeology, it also occurred, albeit less frequently than with commercial archaeological players. This correlates with the above-described trend on text similarity: texts produced in a competitive developer-led archaeology are both more similar and have blocks of text reproduced from previous reports. It is important to underline the dramatic rise of text reproduction by commercial players. One could argue that developer-led archaeology encourages such a practice; even at public and scientific institutions, it leads to a more factory-model archaeology where reports are treated more as forms to be filled in than actually writing a report from the ground up and going through the entire arc of the scientific process.

Lexical complexity: lexical density and use of specialized language

The discipline of archaeology in Belgium and Flanders is often described as being underdeveloped up to the 2000s. The lack of clear archaeological legislation is especially lamented and connected to the poor state of the field (De Clercq et al. 2013; Van Looveren 2014). It is true that up to the 1980s, a lot of sites were lost and excavations happened on an ad hoc basis. This was followed in the 1990s by a period with slightly better legislation for rescue archaeology, where mainly the national government and universities conducted research. During this period also, a special framework for temporary employment by public institutions encouraged more excavation and publication of archaeological data (Van Looveren 2014, 387). Despite many legislative issues and a chronic lack of funding, one could characterize the archaeological practice in Flanders as a publicly organized rescue archaeology: socialist preventive archaeology "light." Up to the mid-2000s, many excavation results were reported in peer-reviewed articles or monographs published by public institutions.

If one looks at the lexical density (Figure 6) of the corpus produced during the so-called difficult period of 1945–2003/2004, one can discern a writing style with varied and rich vocabulary. Whereas the ratio fluctuates up to the 1980s, between the mid-1980s and early 2000s, it remains steady and high. During this period, the index is around 0.8. Linguists would characterize this as very dense and connect it to analytical writing rich in information-carrying words. This is also the period where theoretical concepts are used more explicitly (see above).

Also during the 1960s, the texts of especially public institutions seem to be denser and of higher analytical depth. This is something we also witnessed during our concept mining, indicating a more profound engagement with theory and interpretative frameworks. Interestingly, the late 1950s–1960s is the period when archaeology was temporarily integrated as a separate department in the larger Royal Institute for Cultural Heritage (Koninklijk Instituut voor het Kunstpatrimonium). This large national heritage institute consisted of various departments engaged in diverse types of

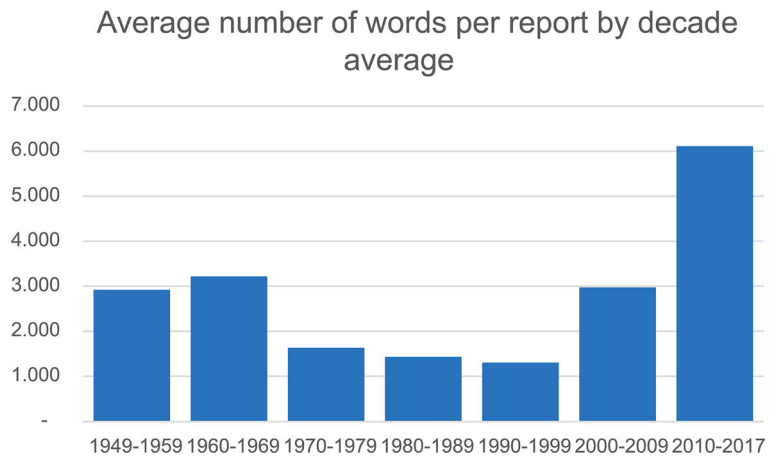

Figure 3. Average length of texts per year.

heritage preservation and art restoration. Within this setting, it received more financial support for fundamental research (*navorsing*), and frequent exchange was encouraged with other departments (e.g. chemists and physicists were included in projects) and academics (Van Looveren 2014, 362–371). It seems that the works produced under these organizational circumstances had more analytical depth.

From the mid-2000s onwards, archaeology became part of the spatial planning process. Importantly, a developer-led type of rescue archaeology became central in the management of cultural heritage. At the start, public heritage institutions and universities were subcontracted by developers. Quickly, private companies became the central archaeological players. A lack of clear guidelines and competition drove prices of fieldwork down. During this period, grey reports become the dominant publication format through which archaeological knowledge was disseminated. Many journal or book series edited by public institutions stopped.

When analyzing the texts produced under a market-driven regime, lexical complexity decreases. When universities and public institutions were subcontracted, the STTR decreased, albeit to an acceptable level. This trend corresponds to the above-described increase in boiler templating by universities and public institutions. Clearly, the quality of the work decreases under a more market-driven archaeology. By the 2010s, when private companies started to define the field, texts become half as dense compared to a decade before. Whereas texts by public players are of a relative quality and fall within the range of what would be labeled as expository texts, reports by companies fall below the threshold and fall within the range of either general prose or spoken language. When looking at the use of theoretical jargon, a similar image presents itself. The decrease in complex vocabulary and widespread boiler templating under a clear capitalist rescue archaeological regime is without a doubt dramatic, indicating that the quality is poorer compared to previous years.

Indeed, up to the early 2000s, less archaeological work was carried out, resulting in a loss of archaeological data. And although more sites might be excavated today, resulting in data gain, the relatively low quality of most texts raises the question of whether a highly competitive and costly developer-led archaeology also leads to an information gain (cf. De Clercq et al. 2013). There might be a proliferation in archaeological fieldwork, but this does not mean the discipline is now more developed in Belgium. To the contrary, quality indicators suggest that the millions of words produced are very much similar and lack complexity. The Flemish heritage institute acknowledged this trend and, in 2018, launched a subsidy framework to encourage synthesis and more profound interpretation of these large amounts of data produced. However, faced with the widespread boiler templating still continuing, one could question if the reports and data on which these studies should be based have sufficient methodological and theoretical depth.

Methodological nationalism

Throughout most of the 19th and 20th centuries, artifacts found on the territory of Belgium were represented as embodying "Belgian" history, and archaeologists would identify themselves as specialized in "national (i.e. Belgian) archaeology." Over the past half-century, however, deeply rooted identity politics have drastically reformulated the rules of the game. Over the course of the 20th century, Walloon regionalists and Flemish national movements challenged Belgian unitarianism, and from the 1970–1980s onwards, Belgium evolved slowly towards a federal state. In this struggle, history played a role in normalizing and deeply rooting regional identities as national containers. Whereas the Middle Ages (i.e. Flemish cities) and World War One became part of the political portfolio of the Flemish nation builders, archaeology received little political attention from the Flemish movement. This lack of attention for archaeology stands in especially sharp contrast to Walloon examples where archaeology was used to buttress regionalism (Van Looveren 2014, 456–457).

The absence of explicit politicization is, however, matched by possible banal national framings. Word frequency analysis (Figure 7) shows that up to the mid-1970s, Belgium was the only framework within which archaeologists geographically assembled the past. Flemish framings of archaeology start in 1975 and become dominant in the mid-1990s. Today, Flanders has become the cardinal geographical framework through which both the archaeological record of prehistory and historical periods are described, while Belgian signifiers have reduced significantly. This shift, starting in the late 1970s, correlates with broader developments in Belgian society producing a social milieu in which Flemish-ness has come more to the foreground.

Figure 4. Results of the cosine similarity test. A) Indicates the similarity between each year and the year before. A similarity of 1 indicates almost total overlap. B) Indicates the similarity between the reports of 2017 and preceding years.

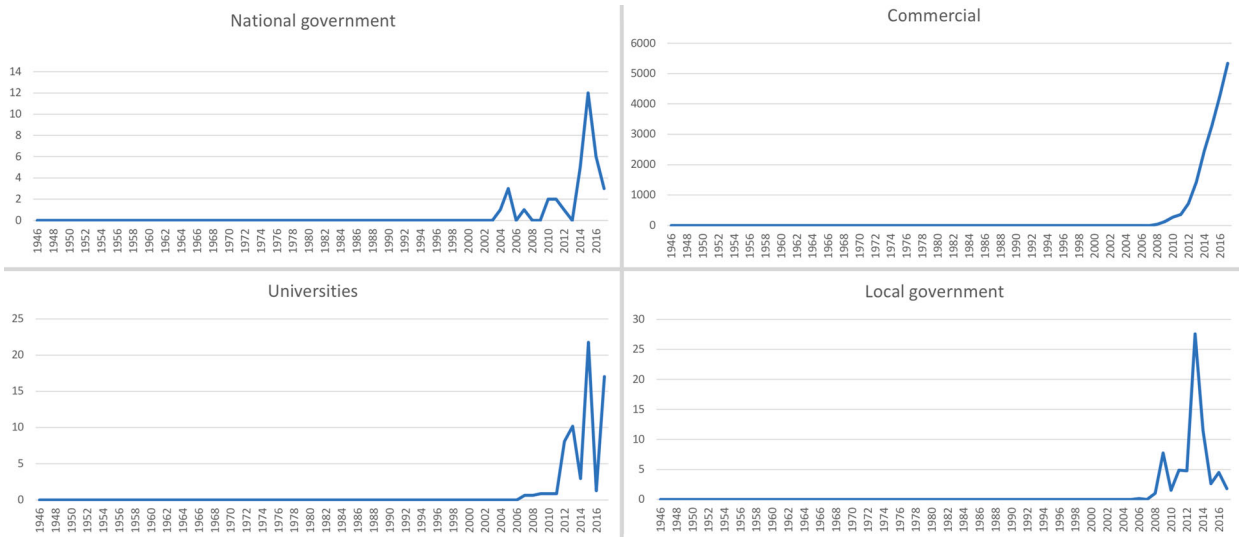

Figure 5. Graphs indicating, per player, how many times in a given year they plagiarized blocks of text of 220 or more characters from previous years. Note: due to the selected parameters, self-plagiarism in which just a couple of words (e.g. site name) are changed could also be detected.

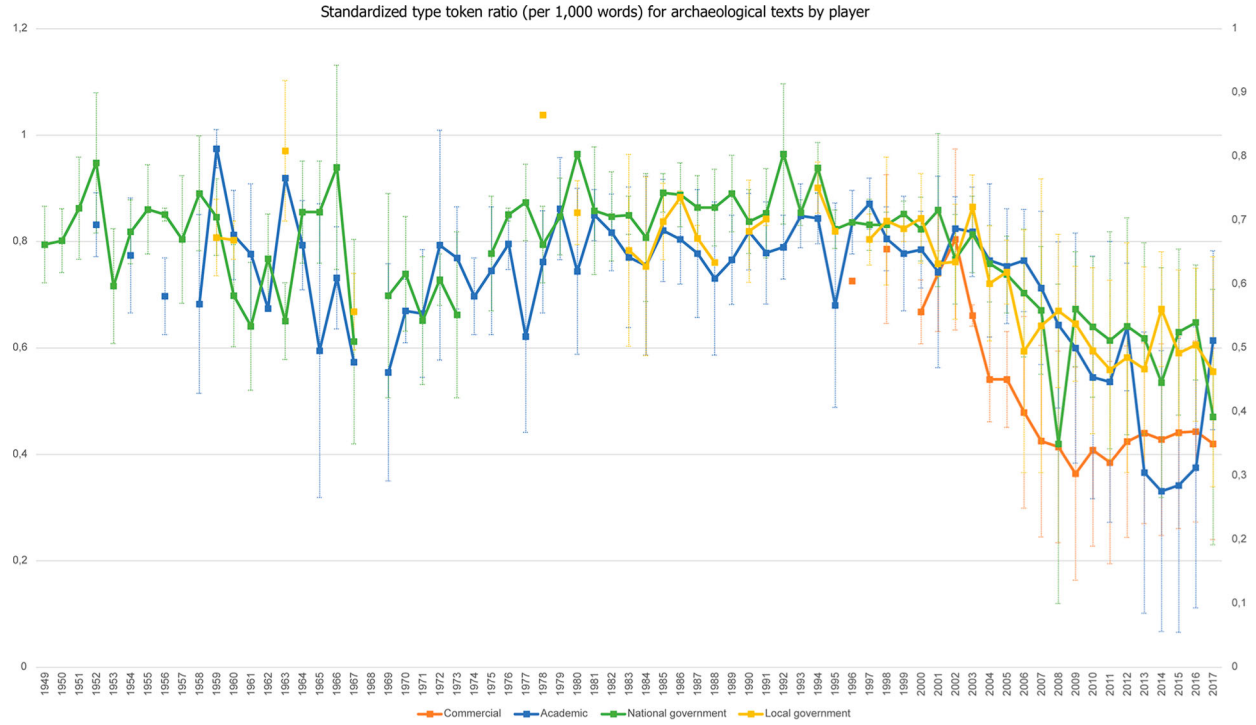

Figure 6. STTR results on the corpus per player. A clear decrease in lexical density can be discerned from 2003–2004 onwards.

Regionalism became an integral part of political and cultural discourse in the public sphere (e.g. media and education), resulting in the formal establishment of a Flemish government and parliament in 1993. Concurrently, key administrative mechanisms in the field of heritage protection have also entangled Flemish archaeologists with the territory of Flanders. From the early 1990s onwards, the Flemish government created legal frameworks, an excavation permit system, and centralized information infrastructures (digitally ordering data within the limits of Flanders) strongly prescribing archaeological praxis. Perspectives from STS teach us that a community subjected to similar technical procedures (Bowker and Star 1999) and archival infrastructures (Derrida 1998) develops a shared identity and reproduces sub-disciplinary boundaries. Clearly, political discourse outside the archaeological field, and very practical bureaucratic changes, have probably encouraged archaeologists to frame material culture in Flemish terms and speak of a "Flemish archaeological past." Further ethnographic and qualitative research is imperative to confirm this indirect proxy.

Europe is largely absent, but steadily increased in the late 1990s and decreased again in the early 2000s. The overall use of geographic signifiers drops significantly from the mid-2000s onwards. This drop correlates with trends discerned above and might present itself as another indication of decreasing interpretative and comparative analysis: sites are merely described and put into a broader geographic context to a lesser degree.

Discussion: Using Text Mining for the Study of Archaeological Thought in Belgium

Over the past decade, important developments within the allied fields of digital linguistics and digital humanities have transformed how historians study changes in knowledge production. Various digital methods were tested on a textual corpus from Belgium (Dutch literature) to evaluate if text mining has the potential to quantitatively chart the impact of a changing context of the archaeologist. Clear trends and developments could be identified using the selected digital tools, underlining the potential of text mining in the study of archaeological knowledge production. Trends that could be mapped correspond with broader theoretical, organizational, and identarian developments.

Clear developments in theoretical traditions could be measured that align with insights provided by qualitative studies. Interestingly, for the Belgian case, more detailed and fine-grained developments could be identified, signaling a strong residual presence of cultural historical concepts and associated empirical discourse. Despite an important rise in processual archaeological frameworks over the past years, the strong presence of the culture-historical tradition in the 1990s is significant. However remarkable it might seem for Anglophone archaeologists, this should not come as a surprise, since continental European archaeologies continue to build and reproduce culture historical themes (Gramsch and Sommer 2011; Hodder 1991) and seemingly engage less with broader theoretical developments in the humanities and social sciences.

The increase and decrease of theoretical discourse at specific periods correlate with the proxies used for assessing research quality and use of geographic signifiers. Periods with significant theoretical engagement were also characterized by a high lexical density and standardized type-token ratio, thus indicating higher scholarly quality. During periods with very low theoretical engagement, the lexical density was also very low. At these moments, a lot of text reuse happened, which was used as a proxy for the use of standardized boiler templates. Importantly, these changes in scholarly quality overlap with changes in the organizational circumstances. During periods where preventive archaeology was led by public institutions, the research quality could be labeled as high—even if the legal framework was insufficient and a lot of sites were being lost to development.

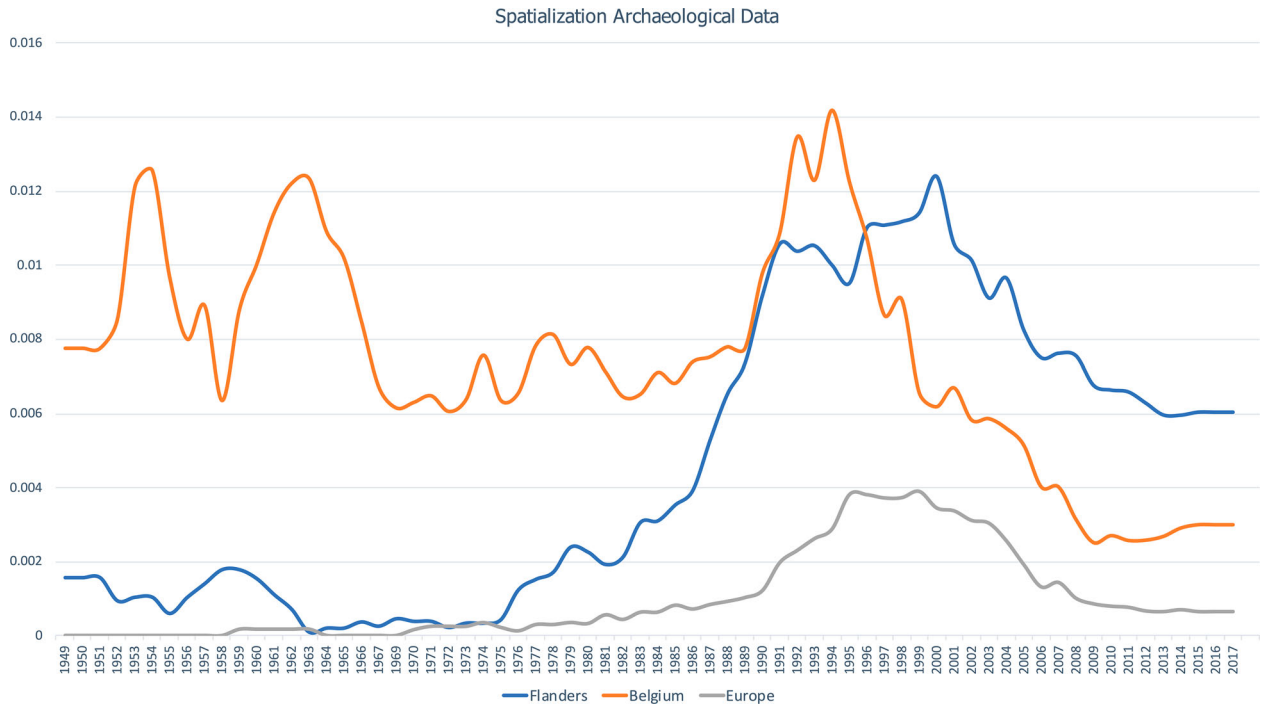

Figure 7. Relative frequency of the used geographic signifiers per year. Since for the early years of the corpus, the data varied per year, we calculated for each year the 3 year average per year (sum of year, previous year, and following year).

From the mid-2000s, more sites were excavated and more data was produced. But as preventive archaeology became governed by a capitalist regime (cf. Kristiansen 2009), the quality of the reports clearly dwindled and intellectual engagement with the archaeological data became minimal.

The almost total disappearance of a theoretical vocabulary can without a doubt be partially related to publication practices in Belgian academia, where university archaeologists are encouraged to publish in English. Interviews show that these publications are rarely consulted by professional archaeologists excavating in Belgium. At the same time, aside from conference proceeding series, there are today few Dutch-language publication venues for both academic and public archaeologists. During the 1990s, most rescue excavations were predominantly published in a suite of journals or special series. Although considerably shorter than excavation reports and not communicating all important details, engaging with an academic debate and contextualizing excavation results or interpretation on a meta-conceptual level was imperative for getting through peer review and reaching the journal's public. Thus, not only the disappearance of universities in the archaeological publication field, but also the changed publication formats (i.e. reports) and venues could explain the strong decrease in academic quality. The key to the solution might lay in encouraging publication beyond the administrative field report or the creation of Dutch language journal series.

At the same time, indirect indications of changing banal nationalist framings of the archaeological past could be identified. These correlate with a changing political landscape where media, digital information, and cultural production outside the archaeological field is increasingly administered by the Flemish state. Furthermore, as noted by Maly (2016), instilling banal nationalism is also a broader ambition of the ruling nationalist Flemish party. Despite the fact that methodological nationalism remains undertheorized in archaeology, based on comparable research in history (Billig 1995; Torsti 2004) and beyond (De Cesari and Rigney 2014; Wimmer and Glick Schiller 2002), we can argue that also in the discipline of archaeology, methodological nationalism is an integrated part of archaeological knowledge production, even if explicit nationalist myths are absent. Although further qualitative research is necessary to develop a more fine-grained understanding of banal nationalism in the Flemish heritage sector, this study still shows a changing use of geographic signifiers in descriptions and contextualizations of the archaeological past. This discursive change is unambiguous in our data and directly correlates with political changes; thus, it is likely the canary in the coal mine indicating changing national frameworks underlying archaeologists' engagement with the past.

Conclusions

Through the data-driven approach presented in this research, changes in archaeological discourse could be traced, analyzed, and contextualized. Trends in the big data analyses correlate with broader changes in the cultural and political field, illustrating the huge potential of digital history methods in the study of archaeological thought and politics. In addition to tracing trends, this study could also quantitatively measure which organizational circumstances in preventive archaeology yield the best research quality: under a public driven regime, there is perhaps less data profit but enormous information profit.

The authors wish to emphasize that this paper constitutes a first step in the use of digital data-driven tools for the study of archaeological knowledge production. We have used only basic standard tools, since these are easily applicable and can be used and tested for other national settings. We also used these basic tools with the clear goal to inspire other archaeologists and scholars of digital humanities to use and test more specialist tools and encourage further computational research on archaeological corpora. The authors will

similarly enrich the research on methodological nationalism by including the maps archaeologists use when contextualizing sites. Experiments with neural networks for the automated classification of images will stand central in this effort. Tests with multi-lingual tools are a logical next step, especially since there is a large French literature and many university archaeologists currently publish in English.

Acknowledgements

The authors wish to thank the Utrecht University Applied Data Science group for the generous funding that made this project possible. The engineers of the Utrecht University Digital Humanities Lab are also recognized for their assistance with the Text Matcher tool. We also wish to thank our informants who made time for us during the winter of 2019–2020 for providing important input regarding the day-to-day working of contract archaeology in Flanders. A previous version of this paper was presented during the research afternoons of the Department of Archaeology of Ghent University and during the yearly lecture of the Alumni Gentse Archeologen (AGAR). We wish to thank the participants of these workshops for their input and often critical remarks. Jean Bourgeois and Jonas van Looveren provided important feedback on previous versions of this paper. Koen van Daele of the Flemish Agency for Immovable Heritage provided access to the digital archive of the Flemish Government. Alexis Wielemans of the same agency is greatly thanked for his assistance during the digitization of the older archaeological texts. His rich institutional memory helped us in developing a large corpus.

Notes on Used Software

The MALLET tool for topic modeling is available at http://mallet.cs.umass.edu/topics.php. The settings that gave the best results were 50 topics, optimize interval 10, and threshold 0.03, with an input of our data split into chunks of 1000 words.

The Voyant tool was used for the concept mining and is available at www.voyant-tools.org. Next to running Voyant Tools on its website, you can run your own server to analyze larger datasets—which is what we did. The suite offers a range of text analysis tools. We used the "trends" tool to calculate and visualize relative word frequencies per year.

The cosine similarity between texts was studied using the Scikit-Learn package, available at https://scikit-learn.org/stable/. To quantify the numbers of times blocks of text were reused or copy-pasted, the Text Matcher tool was used, available at https://github.com/JonathanReeve/text-matcher. Text-matcher allows further tweaking of the matching process by defining the shortest length of matches to include, i.e., threshold (in characters) and Ngrams to match against (in words). The Ngram value that we followed was 3, the preconfigured choice. This means that possible matches were only analysed for plagiarism by the program after 3 matching words at the start and end of every sentence in the studied block were found.

Notes on Contributors

Gertjan Plets (Ph.D. 2013, Ghent University) is an assistant professor in Heritage Studies and Archaeology at the Cultural History section of Utrecht University. His research interests include the politics of heritage preservation and archaeology, the social and cultural dimensions of archaeological knowledge production in Russia, and the evolution of archaeological thought in Europe.

Pim Huijnen (Ph.D. 2011, University of Amsterdam) is an assistant professor in Digital Cultural History at the Department of History and Art History. His research focuses on the use of digital text analysis to investigate the history of science and the history of ideas. He is, in particular, interested in conceptual history and the circulation of knowledge between the scientific, political, and commercial domains and popular culture.

David van Oeveren (M.A. 2020, Utrecht University) is a lecturer at the Department of Technology and Society Studies at Maastricht University. He is interested in applying digital methods to the study of knowledge production and how scientists are shaped by digital methods and infrastructures.

References

Aggarwal, C. 2015. *Data Mining: The Textbook*. Berlin: Springer.

Aitchison, K. 2010. "Grey Literature, Academic Engagement, and Preservation by Understanding." *Archaeologies* 6 (2): 289–300.

Billig, M. 1995. *Banal Nationalism*. London: Sage.

Bowker, G., and S. L. Star. 1999. *Sorting Things Out: Classification and Its Consequences*. Cambridge: MIT Press.

Börjesson, L. 2015. "Grey Literature - Grey Sources? Nuancing the View on Professional Documentation: The Case of Swedish Archaeology." *Journal of Documentation* 71 (6): 1158–82.

Brandsen, A., K. Lambers, S. Verberne, and M. Wansleeben. 2019. "User Requirement Solicitation for an Information Retrieval System Applied to Dutch Grey Literature in the Archaeology Domain." *Journal of Computer Applications in Archaeology* 2 (1): 21–30.

Brown, R. H., and B. Davis-Brown. 1998. "The Making of Memory: The Politics of Archives, Libraries and Museums in the Construction of National Consciousness." *History of the Human Sciences* 11 (4): 17–32.

Carver, M. 2007. "The Profession and the Academy: Common Objectives." In *Archaeology in Ireland: A Vision for the Future*, edited by G. Cooney, 115–26. Dublin: Republic of Ireland Academy.

Champion, T. 1991. "Theoretical Archaeology in Britain." In *Archaeological Theory in Europe: The Last Three Decades*, edited by Ian Hodder, 129–60. London: Routledge.

Crombe, P. 1996. *Epipaleolithicum En Vroeg- En Midden-Mesolithicum in Zandig Vlaanderen : Bijdrage Tot De Studie Van De Typo-Chronologie En De Nederzettingsstructuur*. Gent: RUG.

de Bolla, P. 2013. *The Architecture of Concepts: The Historical Formation of Human Rights*. New York: Fordham University Press.

De Cesari, C., and Ann Rigney. 2014. "Introduction: Beyond Methodological Nationalism." In *Transnational Memory: Circulation, Articulation, Scales*, 1–15. Berlin: De Gruyter.

De Clercq, W. 2009. *Lokale Gemeenschappen in Het Imperium Romanum Transformaties in Rurale Bewoningsstructuur En Materiële Cultuur in de Landschappen van Het Noordelijk Deel van de Civitas Menapiorum. (Provincie Gallia-Belgica, ca. 100 v. Chr. – 400 n. Chr.)*. Gent: Universiteit Gent.

De Clercq, W., M. Bats, J. Bourgeois, P. Crombé, G. De Mulder, J. De Reu, D. Herremans, P. Laloo, L. Lombaert, G. Plets, and J. Sergant. 2013. "Development-Led Archaeology in Flanders: An Overview of Practices and Results in the Period 1990–2010." In *Development-Led Archaeology in Northwest Europe: Proceedings of a Round Table at the University of Leicester 19th–21st November 2009Proceedings of a Round Table at the University of Leicester 19th–21st November 2009*, edited by L. Webley, M. Vander Linden, C. Haselgrove, and R. Bradley, 29–55. Oxford: Oxbow Books.

de Laet, S. 1974. *Prehistorische Kulturen in Het Zuiden Der Lage Landen*. Wetteren: Universa.

de Laet, S., and W. Glasbergen. 1959. *De Voorgeschiedenis Der Lage Landen*. Groningen: J. B. Wolters.

De Mulder, G., and J. Bourgeois. 2017. "On Migrations: Sigfried Jan De Laet (1914-1999): His Role in Belgian Bronze Age Archaeology After the Second World War and the Diffusion of Cultural Characteristics." In *Movement, Exchange and Identity in Europe in the 2nd and 1st Millennia BC: Beyond Frontiers*, edited by A. Lehoërff and M. Talon, 10–20. Oxford: Oxbow Books.

Demoule, J.-P. 2012. "Rescue Archaeology: A European View." *Annual Review of Anthropology* 41: 611–26. https://doi.org/10.1146/annurev-anthro-092611-145854.

Derrida, J. 1998. *Archive Fever: A Freudian Impression*. Chicago: University of Chicago Press.

Diaz-Andreu. 2008. *A World History of Nineteenth-Century Archaeology: Nationalism, Colonialism, and the Past*. Oxford: Oxford University Press.
Dietler, M. 1994. "'Our Ancestors the Gauls': Archaeology, Ethnic Nationalism, and the Manipulation of Celtic Identity in Modern Europe." *American Anthropologist* 96 (3): 584–605.
Evans, T. 2015. "A Reassessment of Archaeological Grey Literature: Semantics and Paradoxes." *Internet Archaeology* 40, https://doi.org/https://doi.org/10.11141/ia.40.6.
Geismar, H. 2015. "Anthropology and Heritage Regimes." *Annual Review of Anthropology* 44: 71–85. https://doi.org/10.1146/annurev-anthro-102214-014217.
Gramsch, A., and U. Sommer. 2011. *A History of Central European Archaeology: Theory, Methods, Politics*, edited by Alexander Gramsch and Ulrike Sommer. Budapest: Archaeologingua.
Guldi, J. 2019. "The Measures of Modernity: The New Quantitative Metrics of Historical Change Over Time and Their Critical Interpretation." *International Journal for History, Culture and Modernity* 7 (1): 899–939.
Hall, D., D. Jurafsky, and C. D. Manning. 2008. "Studying the History of Ideas Using Topic Models." In *EMNLP 2008 - 2008 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference: A Meeting of SIGDAT, a Special Interest Group of the ACL*, 363–71. https://doi.org/10.3115/1613715.1613763.
Halliday, M. 1989. *Spoken and Written Language*. Oxford: Oxford University Press.
Hodder, Ian. 1989. "Writing Archaeology: Site Reports in Context." *Antiquity* 63 (239): 268–74.
Hodder, I. 1991. *Archaeological Theory in Europe: The Last Three Decades*. London: Routledge.
Johansson, V. 2009. "Lexical Diversity and Lexical Density in Speech and Writing: A Developmental Perspective." *Working Papers in Linguistics* 53: 61–79. https://doi.org/10.7820/vli.v01.1.koizumi.
Johnson, M. 2011. "On the Nature of Empiricism in Archaeology." *Journal of the Royal Anthropological Institute* 17: 764–87.
Johnson, M. 2019. *Archaeological Theory: An Introduction*. Oxford: Wiley-Blackwell.
Kim, M., S. A. Crossley, and K. Kyle. 2018. "Lexical Sophistication as a Multidimensional Phenomenon: Relations to Second Language Lexical Proficiency, Development, and Writing Quality." *Modern Language Journal* 102 (1): 120–41. https://doi.org/10.1111/modl.12447.
Kristiansen, K. 2001. "Borders of Ignorance: Research Communities and Language." In *Quo Vadis Archaeologia? Whither European Archaeology in the 21st Century?*, edited by Z. Kobylinski, 38–46. Warsaw: Institute of Archaeology and Ethnology Polish Academy of Sciences.
Kristiansen, K. 2009. "Contract Archaeology in Europe: An Experiment in Diversity." *World Archaeology* 41 (4): 641–48. https://doi.org/10.1080/00438240903371486.
Kristiansen, K. 2012. "Archaeological Communities and Languages." In *The Oxford Handbook of Public Archaeology*, edited by R. Skeates, C. McDavid, and J. Carman, 461–77. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199237821.013.0024.
Lucas, G. 2018. *Writing the Past: Knowledge and Literary Production in Archaeology*. London: Routledge.
Maly, I. 2016. "'Scientific' Nationalism: N-VA and the Discursive Battle for the Flemish Nation." *Nations and Nationalism* 22 (2): 266–86.
Marien, M. E. 1952. *Oud België: Geschiedenis van België van de Eerste Landbouwers tot de Komst van Caesar*. Brussel: De Sikkel.
Meskell, L. 1998. *Archaeology Under Fire: Nationalism, Politics and Heritage in the Eastern Mediterranean and Middle East*. London, United Kingdom: Routledge.
Morretti, F. 2013. *Distant Reading*. London: Verso.
Niklasson, E. 2013. "Appropriate Narratives: Archaeologists, Publics and Stories." In *Appropriate Narratives: Archaeologists, Publics and Stories*, edited by Elisabeth Niklasson and Thomas Meier, 49–86. Budapest: Archaeolingua Alapitvany.
Plets, G. 2016. "Heritage Bureaucracies and the Modern Nation-State: Towards an Ethnography of Archaeological Systems of Government." *Archaeological Dialogues* 23 (2): 193–213.
Plets, G., P. Huijnen, and D. van Oeveren. 2021. "Excavating Flemish Archaeological Texts [Data Set]." Zenodo. 2021. https://doi.org/10.5281/ZENODO.4442172.
Richards, J., D. Tudhope, and A. Vlachidis. 2015. "Text Mining in Archaeology: Extracting Information from Archaeological Reports." In *Mathematics and Archaeology*, edited by J. A. Barcello, and I. Bodganovic, 240–54. Boca Raton: CRC Press. https://doi.org/https://doi.org/10.1201/b18530-15.
Roth, B. 2010. "An Academic Perspective on Grey Literature." *Archaeologies* 6 (2): 337–45.
Seymour, D. 2010. "Sanctioned Inequity and Accessibility Issues in the Grey Literature in the United States." *Archaeologies* 6: 233–69.
Tahmasebi, N., L. Borin, and A. Jatowt. 2018. "Survey of Computational Approaches to Lexical Semantic Change." *Computational Linguistics* 1 (1): 1–55.
Tangherlini, T. 2013. "The Folklore Macroscope: Challenges for a Computational Folkloristics." *Western Folklore* 72 (1): 7–27.
Thoen, H. 1987. *Belgische Kustvlakte in de Romeinse Tijd*. Brussel: Gemeentekrediet.
Thomas, J. 2004. *Archaeology and Modernity. Archaeology and Modernity*. London: Routledge. https://doi.org/10.4324/9780203491119.
Torsti, P. 2004. "History Culture and Banal Nationalim in Post-War Bosnia." *South East European Politics Online* 5 (2–3): 142–57.
Trigger, B. G. 2006. *A History of Archaeological Thought*. Cambridge: Cambridge University Press. https://doi.org/10.1017/cbo9780511813016.
Tys, D. 2004. "De Inrichting van Een Getijdenlandschap. De Problematiek van de Vroegmiddeleeuwse Nederzettingsstructuur En de Aanwezigheid van Terpen in de Kustvlakte: Het Voorbeeld van Leffinge (Gemeente Middelkerke, Prov. West-Vlaanderen)." *Archeologie in Vlaanderen* 8: 257–79.
Van Looveren, J. 2014. *IJdele Hoop? Een Politiek-Institutionele Geschiedenis van Archaeologische Monumentenzorg in België (1830-1991/3)*. Antwerpen: Universiteit Antwerpen.
Vander Linden, M., and L. Webley. 2013. "Introduction: Development-Led Archaeology in Northwest Europe. Frameworks, Practices and Outcomes." In *Development-Led Archaeology in Northwest Europe: Proceedings of a Round Table at the University of Leicester 19th–21st November 2009*, edited by L. Webley, M. V. Linden, C. Haselgrove, and R. Bradley, 1–8. Oxford: Oxbow Books.
Wiedemann, G., M. Lemke, and A. Niekler. 2013. "Postdemokratie Und Neoliberalismus: Zur Nutzung Neoliberaler Argumentationen in Der Bundesrepublik Deutschland 1949-2011; Ein Werkstattbericht." *ZPTh - Zeitfschrift Für Politische Theorie* 4 (1): 99–115.
Willems, W., and M. van den Dries. 2007. "The Origins and Development of Quality Assurance in Archaeology." In *Quality Management in Archaeology*, edited by W. Willems and M. van den Dries, 1–12. Oxford: Oxbow Books.
Wimmer, A., and N. Glick Schiller. 2002. "Methodological Nationalism and Beyond: Nation-State Building, Migration and the Social Sciences." *Global Networks* 2 (4): 301–32.