The Eurocentric Fallacy. A Digital-Historical Approach to the Concepts of 'Modernity', 'Civilization' and 'Europe' (1840–1990)

Joris van Eijnatten and Ruben Ros

Abstract

According to recent literature, the idea of Europe as it developed during the nineteenth and twentieth centuries coincided closely with the concepts of 'civilization' and 'modernity'. This article examines this claim by testing the existence of modernity, civilization and Europe as a conceptual 'trinity' by using digital history techniques. Word frequencies, collocations and word embeddings are employed to analyze four Dutch newspapers (*Algemeen Handelsblad*, *Nieuwe Rotterdamsche Courant*, *De Telegraaf*, *De Leeuwarder Courant*) spanning the period 1840–1990. It transpires that semantic relations between the three elements are hardly visible; in so far as they appear, they hardly constitute a close-knit 'trinity'. Alternating combinations among the three components were more significant than direct connections between all three, while 'the West' was more central than Europe. These findings suggest that popular media like newspapers have a different take on cultural-political concepts than writings by the intellectual elite.

Keywords: civilization, collocations, digital history, digital humanities, Europe, modernity, n-grams, word embeddings

Introduction[1]

This article examines a claim commonly put forward in the literature: that the idea of Europe as it emerged during the early modern period and developed over the nineteenth and twentieth centuries coincided with both 'civilization' and 'modernity'. Like modernity, so the argument runs, civilization was based on the idea of a rupture in time that set the historical present apart from its past.[2] The 'critical elements' in this 'European master narrative', as one recent account typically frames the argument, are a range of civilizational developments that have underlain definitions of modern Europe since at least the nineteenth century: 'the struggle between religion and secularism that encourages tolerance and freedom of conscience; the overthrow of medieval feudalism that paves the way for the rational state; divisive ethnic and tribal antagonisms reconciled within the civic unity of the nation-state'.[3] Or, as another recent account puts it:

> Searching for the roots of what he should call occidental rationalism, Max Weber would famously ask in the early twentieth century, 'what concatenation of circumstances has led to the fact that in the Occident, and here only, cultural phenomena have appeared which – as at least we like to imagine – lie in a direction of development of universal significance and validity?'

'Europe became the home of civilization, and the lack of civilization became the means of demarcating the Other', the authors point out in their overview of nineteenth and twentieth European thought.[4] Civilizational values are modern values, and modern values are European values: this, according to many, has been the intellectual foundation underlying European thought and action from the Enlightenment to the European Union. The aim of this article is to test this claim against specific source material within a specific national context over a larger period of time.

To flesh out the relations between the conceptual trinity of Europe, civilization and modernity, we focus on newspapers. These are relevant to our aim because they represent a broader swathe of opinion than the material usually examined in conceptual history, such as philosophical treatises, history writing and political documents. The 'density' of conceptual understanding in newspaper articles is, generally speaking, much lower than in writings that are dedicated to intellectual reflection

(although newspapers also include in-depth prose). In some respects this is an obvious drawback in tracing concepts over time. However, newspapers represent the general development, if not of public opinion, then at least of the general language use that both feeds into, and draws on, more considered intellectual thought. Moreover, because newspapers are serial in nature, their contents can be traced over larger periods of time, in our case about 150 years.

There is another advantage to using newspapers, which is that they are amenable to being explored by computers if they are digitized. This greatly expands the scale of research and in our case allows us to answer the question of how, exactly, words associated with the concepts under scrutiny arose over time. The article thus attempts to push forward work on conceptual history by employing computational methods. The latter have been increasingly employed in the historical discipline over the past years but digital conceptual history as such is still in its infancy.[5] This article, however, does not aim to develop a new methodology or introduce new digital tools; it seeks to apply existing methods to traditional conceptual history in order to understand how the threefold conceptual triad of Europe, civilization and modernity figured in the public opinion of the past. Historians of concepts have long recognized the value of tracing patterns in word use to understand conceptual change.[6] Digital datasets and computational methods now enable more advanced enquiries into frequencies and distributions that reveal aspects of conceptual change that previously remained undetected.[7]

Conceptual history, whether approached digitally or not, is necessarily concerned with specific linguistic contexts. We have opted to study the Dutch context for three reasons. Firstly, the Netherlands, partly because it is a relatively small country surrounded by powerful neighbours and partly because of its dependence on international trade, has been heavily involved in the post-war creation of what ultimately developed into the European Union. Assuming that Europe will have resonated in this particular country before World War II, the Dutch case offers an excellent opportunity for testing the extent to which the entanglement between Europe, modernity and civilization obtained in a relatively prominent European country. Secondly, in the Netherlands large quantities of digitized, open-access newspapers spanning the nineteenth and twentieth centuries are readily available. Lastly, both authors are native speakers of Dutch.

This article, then, is a case study to test whether the formula: Europe = modernity = civilization does, in fact, occur in Dutch newspapers. We want to ascertain whether newspapers actually represented this triad as a close-knit trinity. Were these concepts as thoroughly intertwined in 'public opinion' as the literature claims them to be? Because Europe is a conceptually variable and imprecise term, we have approached the first member of the 'MCE' trinity (modernity, civilization, Europe, in that order) from the perspective of the other two: modernity and civilization. Our aim is to capture the semantic relations between the various members. To do this, we need first to trace the development of each member of the MCE trinity. From this perspective a digital history approach offers an added benefit. We know from the literature which semantic changes to expect, but we have no grasp of the scale of those changes and little insight into the patterns of word usage that reflect them. This article offers such insight.

In what follows, we first provide a brief account of our methodology (on aspects of which we elaborate in an Appendix). Subsequently we trace word usage concerning modernity and civilization as it develops over time (respectively in the sections 'The Rise of Modernity' and 'The Vagaries of Civilization'). We then examine the conceptual entanglement of both modernity and civilization with Europe ('Modern Europe' and 'Civilized Europe'). Finally, we discuss and contextualize the outcomes of our analysis.

Methodology

Because most historians will still be unfamiliar with digital history methods, a brief account of our methodology is unavoidable. The findings of this article are based on newspaper data obtained from the Dutch National Library (the Koninklijke Bibliotheek in The Hague).[8] We have restricted ourselves to corpora based on four Dutch-language newspapers issued between 1800 and 1990: *Algemeen Handelsblad* (henceforth *AH*, 1828–1970), *Leeuwarder Courant* (*LC*, 1800–1990), *Nieuwe Rotterdamsche Courant* (*NRC*, which has been partly digitized for 1844–1869, 1909–1929, 1970–1990; note that *AH* and *NRC* merged in 1970 as *NRC Handelsblad*) and finally *De Telegraaf* (*TEL*, 1893–1990). The reasons for choosing these specific newspapers are

threefold. Firstly, the newspapers are fairly representative of the Dutch newspaper landscape, in both a temporal and spatial sense.[9] Together, they cover a large part of the nineteenth and twentieth centuries. With the exception of the first decades of the nineteenth century, the overlap between them allows for comparison across both newspapers and time. For the sake of comparison, we have divided, wherever appropriate, the one and a half centuries covered in this article into five periods: 1840–1869 (P1), 1870–1899 (P2), 1900–1939 (P3), 1950–1969 (P4) and 1970–1990 (P5). Table 1 offers some statistics based on this division into periods.[10]

The five blocs (P1–P5) correspond reasonably well with changes in spelling, the huge impact of World War II on the data and the gaps in the *NRC* dataset.

Another reason for selecting these newspapers is that they had a broad national reach, even while each had its own regional basis, certainly in the first half of the period: *LC* had its base in Leeuwarden in the northern Netherlands, *NRC* in Rotterdam, and *TEL* and *AH* largely in Amsterdam. Each newspaper tried to cater to a broader audience but, of course, each also had its specific ideological leanings. *LC* had a Protestant background, while *AH* and *NRC* tended to have a 'liberal' audience, which in the Dutch context basically meant the moneyed class. *AH* and *NRC* merged in 1970; the resulting *NRC Handelsblad* (to which we will continue to refer as *NRC*) became a quality newspaper for the intellectual and political elite with slightly right-wing leanings for much of the period. *TEL*, on the other hand, was a popular, if not populist, daily with a decidedly conservative slant.[11] This brings us to a third selection criterion: in accordance with the literature, we expect right-wing newspapers to show relatively more interest in the MCE trinity.[12] Catholic newspapers would have been a welcome addition in this respect, but they were not included in the sample because they did not contain sufficient data for digital analysis. Wherever possible, we have expressly shown developments per newspaper.

We have adopted the axiom developed by Reinhart Koselleck, that concepts are expressed in words but at the same time are more than words. Words become concepts at the moment when 'the plenitude of politico-social context of meaning and experience in and for which a word is used can be condensed into one word'.[13] While the MCE trinity is thus potentially expressed through a larger variety of words, we

Table 1. Overview of the five periods P1–P5, the newspapers examined, the total number of articles per period and the total number of tokens (i.e. words, including stopwords) per period.

	P1: 1840–1869	P2: 1870–1899	P3: 1900–1939	P4: 1950–1969	P5: 1970–1990
Newspapers	AH, LC, NRC	AH, LC, NRC	AH, LC, NRC, TEL	AH, LC, TEL	LC, NRC, TEL
Articles	101,478	422,420	1,606,998	1,375,525	2,182,986
Tokens	3,894,042	14,775,268	120,921,091	47,989,324	61,419,111

focus on the nouns *europa* ('Europe'), *moderniteit* ('modernity') and *beschaving* ('civilization', supplemented where applicable with *cultuur* for 'culture'), and their associated adjectives.[14]

We apply these highly charged words using three approaches. Firstly, we examine so-called unigrams, which are single words, and n-grams, which are contiguous series of tokens, usually two to five. By exploring their occurrences per newspaper per year, we can map the diachronic occurrence and usage of the various keywords. Bigrams in particular are useful, as they often reveal the words accompanying various adjectives across time. We have employed n-grams in three ways: to map the relative frequency of words preceded by adjectives; to calculate 'conceptual productivity' (i.e. change in the number of unique n-grams per year); and to determine the development of specific n-grams over time.

Secondly, we have determined collocations of the selected keywords. This allows us to examine the probability of particular words co-occurring more often than would be expected by coincidence. Collocations expand the investigation into word relations beyond the level of consecutively ordered words. To measure the proximity of words we have opted for Pointwise Mutual Information (PMI), one of the standards for determining the level of association between two values (in our case: two words). As a rule of thumb we can say that the higher the PMI score, the stronger the relationship between the two words.[15]

The third method we employ, in particular to investigate the entanglement of the members of the MCE trinity, is vector semantics, a method currently used to enquire into the semantic similarity between words. The method helps to solve one of the major conundrums in conceptual history: how to find words referring to a specific concept when those words are unknown, for instance due to changes in the linguistic context. In extension, it has also been employed to study diachronic meaning change, making the method attractive to historians. The method rests on the representation of words as multidimensional vectors or 'word embeddings'; the distance between words in vector space allows one to determine the semantic similarity between them (in what follows, we use the term 'most similar words'). The approach is too technical to explain at length here and we refer the reader to the literature.[16] However, it is important to note that with computational implementations of vector space modelling now easily available, the risk of carelessly employing these off-the-shelf tools becomes higher. For this

reason, we describe the choices in data preparation, model training and model alignment in the Appendix at the end of this article.

The Rise of Modernity

It is well known from the literature that words and concepts related to the 'modern' emerged in the Middle Ages, when *modernus* figured as a counterpart for the morally, culturally and theologically superior *antiquus*.[17] The Renaissance saw a more emphatic integration of temporality into the concept, giving rise to a positive appreciation of the modern in the context of the *Querelles des Anciens et Modernes*. Around the turn of the nineteenth century two interpretations emerged: modernity as an aesthetic category and modernity as a historical phase. The concept thus acquired a historical connotation, in particular in connection with the notion of progress. Koselleck theorized that it turned into a 'collective singular' by embodying the various 'progresses' in a unitary idea of Progress. In this way, steam engines, weaponry, imperialism, cubist art and atonal music could all be encapsulated in the concept of modernity, including in western modernity as the expression of a civilization superior to all others. Indeed, the concept of civilization now became intrinsically bound up with modernity, so that 'modern civilization' represented the pinnacle of man's achievements. 'The modern' thus gained a spatial connotation in addition to the temporal one.[18]

We find this link between modernity and civilization in the newspapers we have examined. Some examples may help to illustrate this:

– the men of the reaction worked for six years trying to mould France according to their limited perspective; they were intent on destroying *modern civilization* (*AH* 1878)
– thus *modern civilization* enables the finest fruits to ripen, including in the field of international law (*LC* 1880)
– which uncivilized, savage nation would we be able to bring under the enlightening influence of *modern civilization* by irradiating it with the searchlights of the military? (*AH* 1895)
– that in the interest of world peace and *modern civilization* the final moment has come to take up the burden of international responsibility (*LC* 1904)[19]

How did such understandings of 'the modern' develop in Dutch newspapers over the nineteenth and twentieth centuries? In the following we will look successively at bigram frequencies, bigram productivity and most similar words.

Bigrams. Bigram frequencies with *modern** as the qualifier appear in *LC* as early as the 1760s.[20] They rise gradually only after the mid-nineteenth century, increasing steadily until about 1970 (Fig. 1).

The distortions shown by *TEL* are mostly the result of its dubious role during World War II. The newspaper ceased to function as an independent medium under the occupation and was prohibited from appearing between 1945 and 1949. Between 1800 and 1840 *modern** appears as an adjective proper, mostly in the context of knowledge ('books', 'sciences') as well as popular aesthetic ('taste', 'objects'). Houses are advertised as having 'modern rooms' and awards are given to artists who create 'modern statues'. Interestingly, popular combinations in *LC* in P1 (1840–1869) and P2 (1870–1899) are theological (Table 2). At
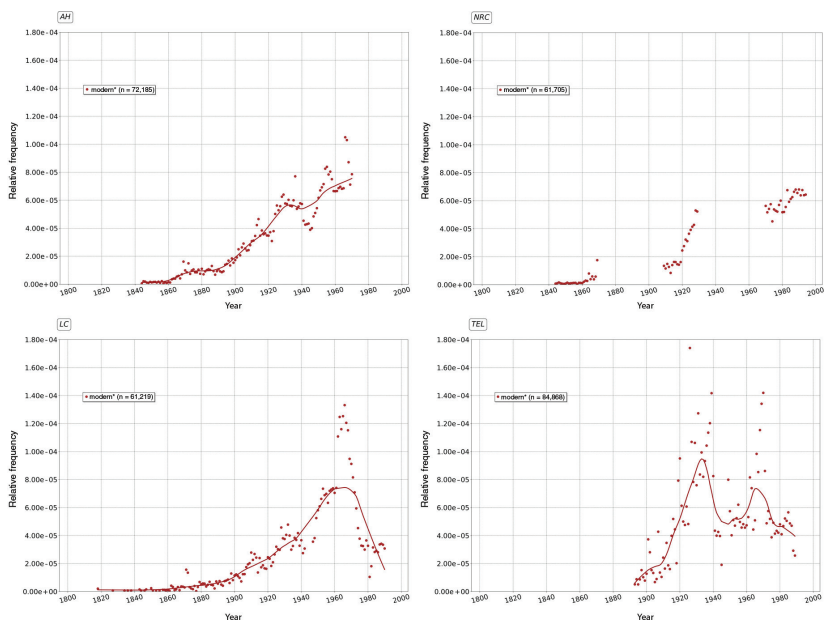

Figure 1. Bigrams of modern* *as adjective, in articles in* AH, NRC, LC *and* TEL *between 1800 and 1990; only words of more than three characters have been taken into account.*

Table 2. Relative frequencies of bigrams with modern* *as their qualifying adjective, in* LC *and* TEL *for periods P1 and P2.*

Bigram '*modern**'		P1: 1840–1869	P2: 1870–1899	
		LC	LC	TEL
richting	school (of thought)	3.24e-05	1.05e-04	2.93e-05
theologie	theology	2.83e-05	7.16e-05	
taal	language	2.43e-05	1.11e-04	1.14e-04
maatschappij	society	1.21e-05	4.96e-05	1.09e-05
schilderij	painting	8.09e-06	5.51e-06	1.63e-04
predikant	preacher	8.09e-06	6.94e-05	2.17e-05
beginselen	principles	8.09e-06		
staat	state	8.09e-06	3.53e-05	2.17e-05
kleding	clothing	4.05e-06		
begrip	concept	4.05e-06	1.10e-05	

the top of the list we can find references to both 'modern theology' and the so-called 'modern school', an influential modernist orientation in progressive Protestant thought. This is closely followed by 'modern language(s)', which remains prominent in all periods.

A second batch of common bigrams in this period refers to 'modern society', 'state' and 'school', a third to 'paintings' and 'music', and a fourth to 'modern concepts'. A similar emphasis on society, art forms and knowledge emerges also in the other newspapers, as these examples show:

- the professional library of Buonaparte. These are mostly *modern* books and publications of the costliest workmanship. (*LC* 1818)
- by moulding [them] to be useful members of society, through a polite education in *modern* languages and other scientific accomplishments (*AH* 1841)
- to resist the supremacy exercised by the state, often with a heavy sceptre over all denominations, according to the *modern* understanding (*NRC* 1849)
- besides mathematics and navigation, old and *modern* languages will also be taught (*NRC* 1855)
- as if the *modern* state had already triumphed irrevocably over the antique one (*NRC* 1868)
- that our whole *modern* state is a creation of unbelief and revolution (*AH* 1868)

–   this demonstrates that fortunately we are not running behind in all aspects of *modern* interaction (*AH* 1904)
–   to unite the conservation of the old with the demands of the *modern* traffic (*AH* 1910)[21]

Between 1900 and 1939 (P3) 'modern times' tops the list in *LC*; bigrams like 'modern life', 'intercourse', 'people' and 'woman' similarly refer to modernity as a feature of a consumer society that offered its members a modern way of life.[22] *LC* mentions how in The Hague, the 'typical vegetable stands have disappeared, and they have been replaced by the commotion of modern traffic'. Another group of bigrams uses different terms for trade unions to identify these as modern. By far the highest score in *LC* in the post-war periods (P4: 1950–1969 and P5: 1970–1990) is for 'music', followed at a distance by 'art', 'chamber music' and 'song'. The patterns in *AH*, *NRC* and *Tel* are not dissimilar.

Productivity. While the bigram productivity for *modern** remained low between 1840 and 1869 (P1), it increased substantially in later years (Fig. 2).
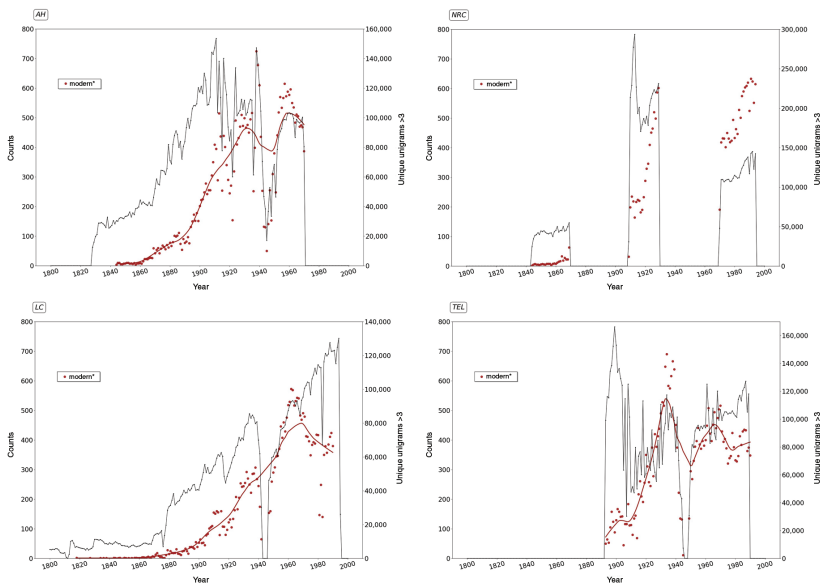

Figure 2. Bigram productivity for modern**, in articles in* AH*,* NRC*,* LC *and* TEL *between 1800 and 1990; only words of more than three characters have been taken into account.*

In other words, 'modern' was used increasingly to qualify a wide range of socio-cultural phenomena. The general pattern is a gradual rise in unique bigrams from around 1860 through to about 1960 or 1970, after which followed a slight decline. *TEL* again shows the distortions resulting from its role in World War II, while *NRC* is difficult to gauge because of the missing material. We can set the productivity off against the total number of unique words (unigrams) in any given year (as in Fig. 2); the latter signifies the richness of the vocabulary in that year. Especially the first decades after World War II seems to have been very productive in generating unique bigrams related to modernity, in particular *NRC*; the odd one out is *TEL*.

New bigrams emerge as culture and society evolve. The top five new bigrams for each period in *TEL* illustrate this:

- 1893–1899: modern painting, modern languages, modern art, modern orientation, modern times
- 1900–1939: modern times (in English), modern union, modern trade union, modern architecture, modern (art of) painting
- 1950–1969: modern countryside, modern gramophone music, modern record company, modern approach, modern millie (a musical)
- 1970–1990: modern talking (in English), modern communication technology, modern love (in English), modern frontier crossing, modern glass art

Vectors. Long-term changes in the semantic context of the adjective 'modern' become clear from diachronic word embeddings. To demonstrate this, we created vector models for P1 through P5 to determine the similarity between words, and then chose the top twenty words most similar to 'modern' in each model. Subsequently, we selected those words that appear in multiple models, ranking them on the basis of their similarity to 'modern' (that is, their 'distance in vector space').[23] The resulting visualization shows how the meaning of 'modern' changes over the course of 150 years (Fig. 3).

It is striking how several words, such as 'new(er)', 'present-day' and 'classical' remain dominant. This does not mean that nothing changes. In P1 'modern' mostly indicates the 'non-past', where temporal distinctions result from the use of words like 'antiquity' and 'classical'. Such distinctions were mostly raised in the context of art, with

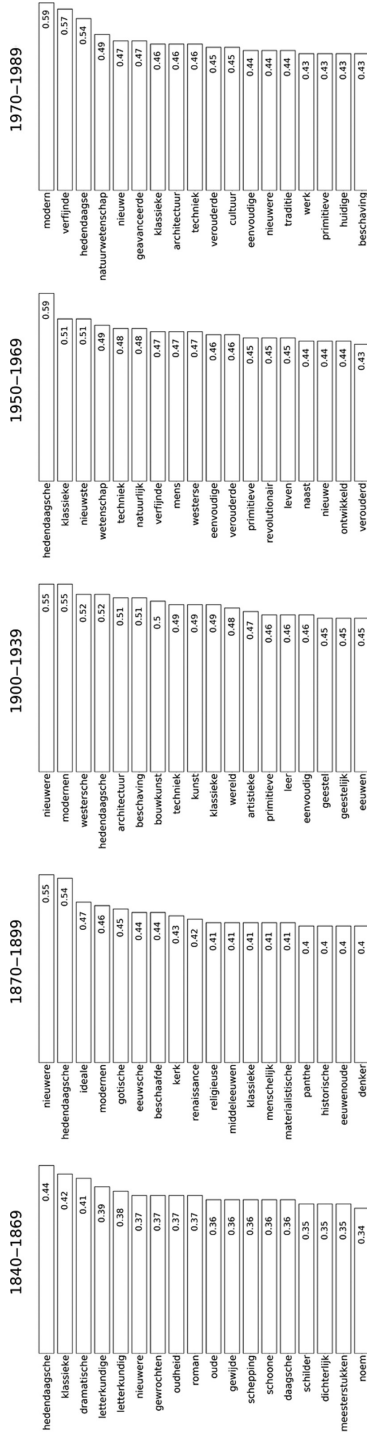| 1840–1869		1870–1899		1900–1939		1950–1969		1970–1989	
hedendaagsche	0.44	nieuwere	0.55	nieuwere	0.55	hedendaagsche	0.59	modern	0.59
klassieke	0.42	hedendaagsche	0.54	modernen	0.55	klassieke	0.51	verfijnde	0.57
dramatische	0.41	ideale	0.47	westersche	0.52	nieuwste	0.51	hedendaagse	0.54
letterkundige	0.39	modernen	0.46	hedendaagsche	0.52	wetenschap	0.49	natuurwetenschap	0.49
letterkundig	0.38	gotische	0.45	architectuur	0.51	techniek	0.48	nieuwe	0.47
nieuwere	0.37	eeuwsche	0.44	beschaving	0.51	natuurlijk	0.48	geavanceerde	0.47
gewrochten	0.37	beschaafde	0.44	bouwkunst	0.5	verfijnde	0.47	klassieke	0.46
oudheid	0.37	kerk	0.43	techniek	0.49	mens	0.47	architectuur	0.46
roman	0.37	renaissance	0.42	kunst	0.49	westerse	0.47	techniek	0.46
oude	0.36	religieuse	0.41	klassieke	0.49	eenvoudige	0.46	verouderde	0.45
gewijde	0.36	middeleeuwen	0.41	wereld	0.48	verouderde	0.46	cultuur	0.45
schepping	0.36	klassieke	0.41	artistieke	0.47	primitieve	0.45	eenvoudige	0.44
schoone	0.36	menschelijk	0.41	primitieve	0.46	revolutionair	0.45	nieuwere	0.44
daagsche	0.36	materialistische	0.41	leer	0.46	leven	0.45	traditie	0.44
schilder	0.35	panthe	0.4	eenvoudig	0.46	naast	0.44	werk	0.43
dichterlijk	0.35	historische	0.4	geestel	0.45	nieuwe	0.44	primitieve	0.43
meesterstukken	0.35	eeuwenoude	0.4	geestelijk	0.45	ontwikkeld	0.44	huidige	0.43
noem	0.34	denker	0.4	eeuwen	0.45	verouderd	0.43	beschaving	0.43

Figure 3. Words most similar to the adjective 'modern' in periods P1–P5. Because multiple models were trained on each period, the figure includes only the most similar words that are present in each model.

reference to, for instance, 'novel', 'fine' (as in fine arts), 'painter' and 'masterpieces'.[24]

This usage of 'modern' as a temporal category spills over into P2, as Fig. 3 shows. But a closer look reveals that other temporal categories are now used as well. Instead of an exclusive focus on the Classical Age, terms that now show up are 'renaissance', 'gothic' and 'Middle Ages'. Together with 'age-old' and 'antiquated', this likely indicates a tonal shift from modernity as a condition of the present ('not old') to modernity as a stage in history ('modern times'). This altered appreciation of modernity is also reflected in specific debates, for example on religion ('church', 'materialist'). Elements from P1 and P2 emerge in P3. The aesthetic modernism of the twenties and thirties is reflected in 'architecture'. At the same time, the normative conception of modernity as fundamentally progressive is pushed to new heights with 'primitive', 'spiritual' and 'refined'. The postwar decades (P4) include many of the elements we have seen in earlier phases. Modern as civilized, modern as a cultural property and modernity as a phase in history all appear in P4 and P5. Specific to these two periods is the more scientific and/or technological conception of the modern, evident from such words as 'technology', 'advanced' and '(natural) science'.[25]

Our observations show that modernity's spatial connotations appear to centre not so much around 'Europe' as around 'the West'. It is the West that is modern, as in: 'Our photographs indicate in a striking manner the difference between modern and Eastern [i.e. Asian] Bangkok'.[26] This particular connection has received ample attention by German scholars of *Begriffsgeschichte*.[27] The authors of a recent work on the topic mention the 'general spatialization of political thought and a reconfiguration of global mental maps that took place between 1780 and 1830'. They claim that early in the nineteenth century the concept of the West was temporalized and politicized, resulting in stronger connections with such notions as progress, modernity and civilization.[28] Additionally, the twentieth century, and especially the postwar period, saw an increasing dominance of Anglo-American models of modernity, resulting in a recalibration of the link between the West and the modern. The list of words in P4 and P5 in Fig. 3 seem to reflect this. Of course, whether these observations on the link between the West, modernity and civilization truly obtain in a longitudinal, big data analysis of Dutch newspapers remains to be proven.

The Vagaries of Civilization

The antecedents of civilization as a concept lie in the eighteenth century. In Dutch, 'civilization' is often translated as *beschaving*, although a case can be made for translating *beschaving* as 'politeness'. While the verbal form goes back to the Middle Ages (when it could mean robbing someone), the nounal form was used in the seventeenth century as a synonym for the French *civilité* and *politesse*. In the later eighteenth century, *beschaving* became a future-oriented, processual word denoting increasing refinement, what Koselleck called a *Bewegungsbegriff* or 'concept of movement'. Hence it was often used interchangeably with *verlichting*, the Dutch word for Enlightenment.[29] Later, in the nineteenth century, civilization became increasingly intertwined with progress and spatial categories such as the nation, the 'West' and Europe.[30] Again, in the following we will successively look at bigram frequencies, bigram productivity and most similar words.

Bigrams. Two words need to be examined, the noun *beschaving* and the adjective *beschaafd*, but since civilization shows semantic overlap with culture, in this section we will also examine *cultuur* and *cultureel*. The pattern for *beschaving** is markedly different from that of *modern** (Fig. 4).

The absolute number of occurrences is substantially lower, while the pattern shows a very clear rise between 1880 and 1940, followed by a decline in the post-war period. The latter was to be expected, since the word went out of fashion after the 1960s. Interestingly, however, *NRC* seems to show an increase in word usage after about 1975 (mirrored to a lesser extent by *TEL*). *NRC* also has a relatively high score in comparison to the other newspapers. The pattern is corroborated by the adjective 'civilized' (Fig. 5).

The top seven bigrams for *beschaving* in *LC* appear to reflect the Western European bias of the term (Table 3).

From P2 through P5 civilization is qualified as 'European', 'Christian', 'western', 'modern' and 'our'. 'True' drops out of the top twenty in P3, 'universal' in P4; 'industrial' emerges in P4, 'extra-terrestrial' in P5. The bigram 'our civilization' generally refers to western (Greco-Roman) civilization in general and Dutch civilization in particular, as these passages from *NRC* illustrate:[31]
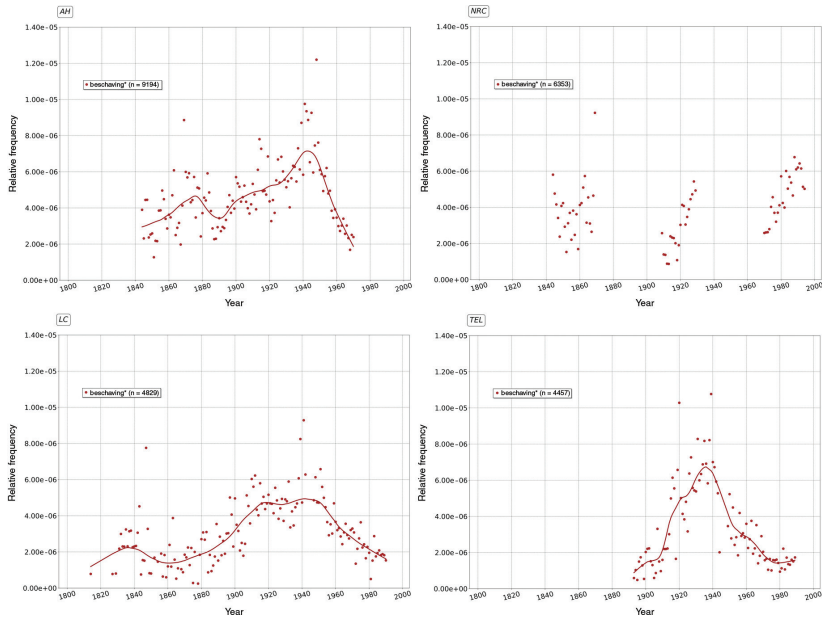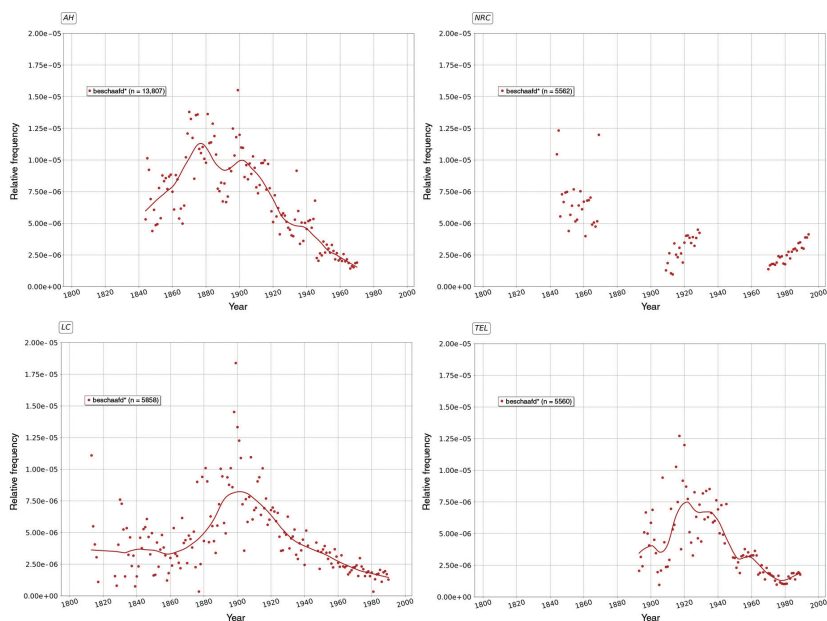
Figure 4. Bigrams of beschaving* *as noun, in articles in* AH*, NRC, LC and* TEL *between 1800 and 1990; only words of more than three characters have been taken into account.*

– Who can prove, prove, that *our civilization* will regress if Greek were abolished? (*NRC* 1924)
– The same [applies] to *our* [word for] *beschaving*: it has little in common with 'civilization', and is actually closer to the French-English *politesse* (*NRC* 1924)
– He recalled the influence which Italian culture, Italian art has had on *our civilization*, on our artists (*NRC* 1924)
– What would happen to *our civilization*, taking this word in its deepest sense, if no other limit were applied to the pursuit of profit than that offered by ordinary criminal law? (*NRC* 1924)

'Dutch' civilization scores high from P3 onwards, but so does 'human' civilization. The other newspapers bear out this pattern. Interesting is the notion of 'white' civilization, which occurs throughout the twentieth century but scores especially high in P4 and P5 (it does not occur in Table 3). The term is sometimes used as a contrast to North America ('redskins'), Africa ('tribes') and Australia ('primitive

Figure 5. Bigrams of beschaafd* *as adjective, in articles in* AH*, NRC, LC and* TEL *between 1800 and 1990; only words of more than three characters have been taken into account.*

natives'), but often simply refers to the dominant culture of colonial European settlers. 'The existence of a white civilization in South Africa depends on the development of Kaffirs in boer municipalities with local self-government'.[32] 'Planters in the tropics who live far from white civilization, can hear the beat of the empire's heart'.[33]

The top bigrams for *beschaafd** for P2–P5 relate to the 'world', 'countries', 'nations' and 'peoples' (Table 4).

This pattern is more or less consistent for each of the three periods. 'Civilized' or 'polite Dutch' crops up, but this is exclusively a truncation of the trigram 'universal polite Dutch', a commonly used term to denote standard Dutch (*algemeen beschaafd Nederlands* or *ABN*). In *LC* the same pattern emerges, with a stronger emphasis on polite middle-class society, a bourgeoisie maintaining cultured 'circles'. In P5 'polite applause' emerges, presumably referring to theatre life in Leeuwarden.

The token *cultuur**, used as a substantive, rises quite spectacularly in the decades after 1880, peaking in the interbellum (Fig. 6).

Table 3. *Relative frequencies of the top bigrams that include the noun* beschaving* *in LC over periods P1–P5, sorted for P1.*

Bigram 'civilization'		P1: 1840–1869	P2: 1870–1899	P3: 1900–1939	P4: 1950–1969	P5: 1970–1990
		LC	*LC*	*LC*	*LC*	*LC*
algemene	universal	2.83e-05	5.51e-06	1.09e-05		
europese	european	2.43e-05	3.86e-05	6.12e-05	2.79e-05	2.35e-05
zedelijke	moral	2.02e-05				
ware	true	2.02e-05	1.54e-05	1.73e-05		
godsdienstige	religious	1.62e-05				
christelijke	christian	1.21e-05	1.21e-05	2.82e-05	3.35e-05	3.74e-05
hoge	high	1.21e-05				
hedendaagse	present-day	1.21e-05	1.43e-05			
geestelijke	spiritual	1.21e-05				
alle	all	8.09e-06		1.17e-05	4.50e-06	
klassieke	classical	8.09e-06				
waarachtige	genuine	8.09e-06	4.41e-06			
griekse	greek	8.09e-06			1.18e-05	6.53e-06
onze	our	8.09e-06	2.53e-05	9.23e-05	1.03e-04	7.536e-05
burgerlijke	civic, 'middle-class'	4.05e-06				
fijne	fine	4.05e-06	1.10e-05	8.06e-06		
eerste	first	4.05e-06				
echte	real	4.05e-06				

Table 3 (continued)

Bigram 'civilization'		P1: 1840–1869	P2: 1870–1899	P3: 1900–1939	P4: 1950–1969	P5: 1970–1990
		LC	LC	LC	LC	LC
enige	some	4.05e-06				
antieke	antique					
grote	big		7.71e-06			
tegenwoordige	contemporary					
huidige	current					4.79e-06
nederlandse	dutch			2.30e-05	1.05e-05	4.35e-06
buitenaardse	extraterrestrial					6.53e-06
franse	french			1.21e-05		
duitse	german			1.21e-05		
meerdere	greater					
hoge	high		2.97e-05	4.03e-05	1.30e-05	5.66e-06
menselijke	human		4.41e-06	1.53e-05	2.30e-05	4.00e-05

Table 4. Relative frequencies of the top bigrams that include the adjective beschaafd* in LC and TEL over periods P2–P5, sorted for P5.

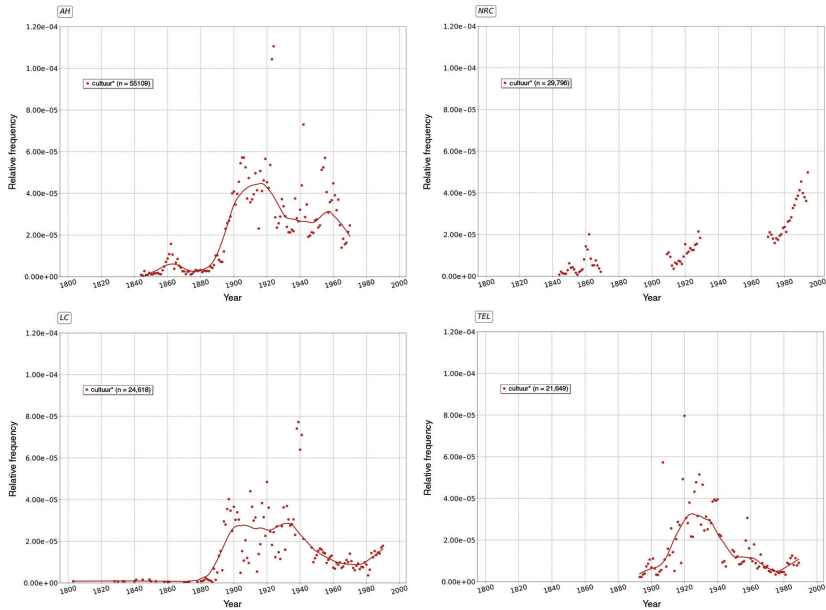| Bigram 'civilized'		P2: 1870–1899		P3: 1900–1939		P4: 1950–1969		P5: 1970–1990	
		LC	TEL	LC	TEL	LC	TEL	LC	TEL
wereld	world	3.06e-04	1.66e-04	2.98e-04	2.48e-04	8.80e-05	8.42e-05	7.18e-05	7.24e-05
landen	countries	7.71e-05	4.44e-05	8.34e-05	7.83e-05	3.22e-05	5.40e-05	2.78e-05	1.79e-05
mensen	people	2.75e-05	1.30e-05	2.90e-05	1.36e-05	1.67e-05	2.12e-05	1.09e-05	1.05e-05
mens	human being	1.87e-05		2.30e-05	1.42e-05	1.98e-05	1.38e-05	8.27e-06	7.33e-06
manier	manner	1.87e-05				5.58e-06	5.82e-06	7.83e-06	5.95e-06
wijze	way					1.05e-05	6.88e-06	3.48e-06	5.95e-06
indruk	impression					6.81e-06	5.82e-06		5.50e-06
stem	voice				7.80e-06	1.24e-05	1.32e-05	3.05e-06	4.58e-06
vrouw	woman	2.86e-05	1.84e-05	1.61e-05	8.96e-06		1.06e-05		4.12e-06
natie	nation	2.42e-05	2.17e-05	1.21e-05	1.36e-05	4.96e-06	4.76e-06		2.75e-06
volkeren	peoples	1.98e-05		3.22e-05	1.53e-05	5.58e-06	4.76e-06	3.48e-06	1.83e-06
naties	nations	5.62e-05	2.06e-05	2.30e-05	3.18e-05	4.34e-06	7.94e-06		1.37e-06
volken	peoples	6.06e-05	2.60e-05	4.96e-05	4.56e-05	4.96e-06	7.41e-06		1.37e-06
jongeman	young man						4.76e-06		1.37e-06
staten	states	1.98e-05	1.73e-05	3.10e-05	2.20e-05		7.41e-06		
applaus	applause							3.92e-06	
eeuw	century								
kringen	circles	2.53e-05	6.50e-06	1.78e-05	8.38e-06				
omgangstaal	colloquial language					6.19e-06			
nederlands	dutch			8.86e-06					

Figure 6. Bigrams of cultuur* *as noun, in articles in* AH, NRC, LC *and* TEL *between 1800 and 1990; only words of more than three characters have been taken into account.*

After the war culture appears at a much lower level, with an upward turn in the 1980s. In *NRC* the term points towards two well-known, general meanings of culture in Dutch. In Table 5, culture as 'cultivation' turns up as, for instance, free culture, the political alternative to the forced cultivation of crops; likewise, 'new' and '(East) Indian' culture have high scores in P3.

Colonial and trade interests account for extensive use of the term. In the earlier period, contractions such as 'sugar culture' are important, while 'flower bulb culture' emerges in P5 (not in the table). By contrast, 'Dutch' culture, in the sense of a collectively shared set of habits and traditions, occurs throughout the twentieth century but especially often in P5. The same applies to European culture. Early examples of the latter include: 'Of course, there is also Japan, and is there a more fertile field for extracting appealing myths than the sunny land of graceful geishas, before it was punished with an invasion of European culture?' 'European culture has to play its last trump now. It will be lost forever if the massacre were to begin again'.[34] 'Western' culture is used in a similar sense. 'After all, since the principle of European education has

Table 5. Relative frequencies of the top bigrams that include the noun cultuur* *in* NRC *over periods P1, P3 and P5, sorted for P3.*

Bigram 'culture'		P1: 1840–1869	P3: 1900–1939	P5: 1970–1990
		NRC	*NRC*	*NRC*
rubber	rubber		3.18e-04	
javasche	javan		1.05e-04	
suiker	sugar	1.52e-05	8.82e-05	
thee	tea	8.05e-06	8.15e-05	
deli	deli (commercial company)		6.07e-05	
koffie	coffee	3.76e-05	5.81e-05	
nieuwe	new		5.41e-05	4.55e-05
onze	our		3.94e-05	4.47e-04
eigen	one's own	1.25e-05	3.48e-05	2.85e-04
koloniale	colonial		3.06e-05	
insulinde	insulinde		2.82e-05	
moderne	modern		2.64e-05	
nederlandse	dutch		2.26e-05	3.43e-04
oude	old		2.03e-05	
intensieve	intensive		2.01e-05	
westerse	western		1.84e-05	2.24e-04
indische	east indian		1.82e-05	
franse	french		1.45e-05	1.03e-04
amerikaanse	american			6.23e-05
arabische	arabian			4.88e-05
burgerlijke	bourgeois, 'middle-class'			5.77e-05
chinese	chinese			5.31e-05
verpligte	compulsory	1.34e-05		
europese	european			1.69e-04
gedwongen	forced	1.70e-05		
vrije	free	2.65e-04		

been carried through with regard to natives (since 1823), Western culture has indeed been opened up to them, but that has also given them good reason to demand a status equal to that of Europeans'.[35]
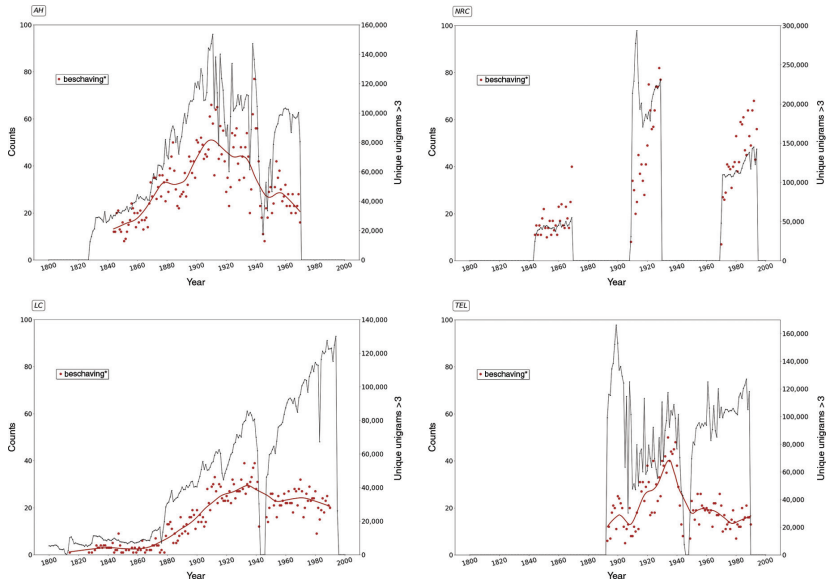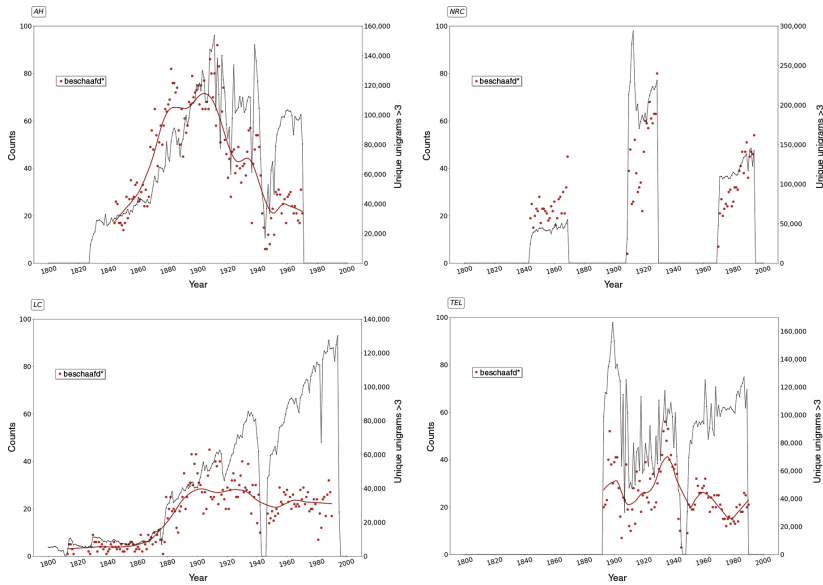
Figure 7. Bigram productivity for beschaving*, *in articles in* AH*,* NRC*,* LC *and* TEL *between 1800 and 1990; only words of more than three characters have been taken into account.*

Productivity. Civilization as *beschaving* begins to generate an increasing number of unique bigrams after about 1860 (in *AH* and *LC*) until it reaches higher levels in the first half of the twentieth century (Fig. 7).

In the second half of that century the productivity of *beschaving* remains comparatively high, but it decreases in relation to the total number of tokens. The single exception is *NRC*, which in the 1980s has the highest variety of bigrams. Examples of novel bigrams for each period (top 5) in *AH* include:

- P1: general civilization/politeness, european civilization, high (level of) civilization, christian civilization, higher (degree of) civilization
- P2: high civilization, greater civilization, fine civilization, more civilization, contemporary civilization
- P3: dutch civilization, one's own civilization, inner civilization, human civilization, big/great civilization
- P4: western civilization, greek civilization, western european civilization, other civilization, white civilization
- P5: extraterrestrial civilization, industrial civilization, whole civilization, current civilization, technical civilization
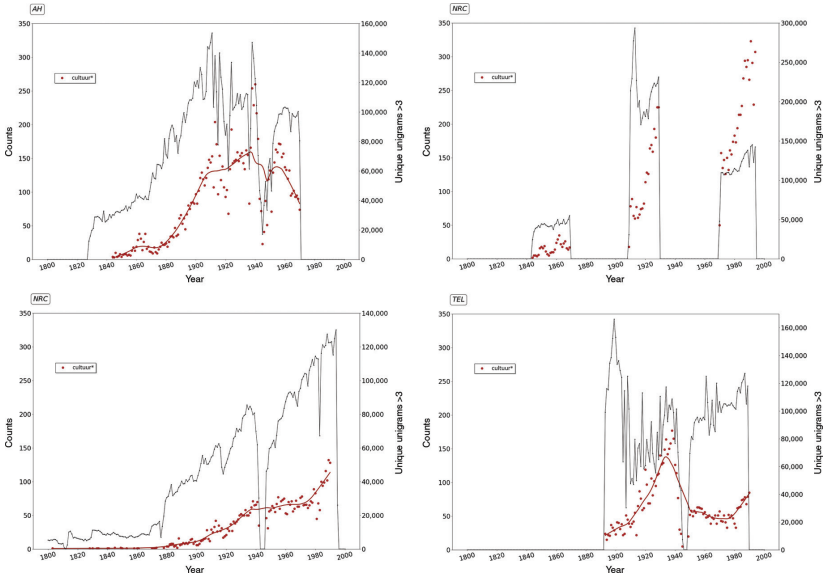
Figure 8. Bigram productivity for beschaafd*, *in articles in* AH*, *NRC*, *LC* and *TEL* between 1800 and 1990; only words of more than three characters have been taken into account.*

The bigram productivity for *beschaafd** is less pronounced in *NRC*, but higher than for other newspapers. This would seem to indicate either more interest in a conservative concept or greater rhetorical versatility, or both (Fig. 8).

Top new bigrams per period in *LC* are, by way of example, 'civilized world' (P1), 'civilized country' (P2), 'civilized/polite language use' (P3), 'civilized person' (P4) and 'polite applause' (P5). The most spectacular rise and fall in bigram productivity occurs in *AH*, which peaks around 1900 and then in 1970 drops to the same level as around 1840.

Culture as *cultuur** shows a slightly different pattern, particularly in P5, which marks a pronounced upward turn, if we see *NRC* as a continuation of *AH*. Cultural as *cultureel* grows steadily throughout the century from 1900 to 1990. The exception seems to be *AH*, which, as in the case of *cultuur**, interestingly shows a decline between 1955 and 1970. Again, in absolute terms *NRC* demonstrates the highest productivity (Fig. 9 and Fig. 10).

Examples of novel bigrams for each period (the top 5) in *NRC* are:
  – P1: free culture, private culture, coffee culture, indigo culture, enforced culture

Figure 9. Bigram productivity for cultuur*, *in articles in* AH*, *NRC*, *LC* and *TEL* between 1800 and 1990; only words of more than three characters have been taken into account.*
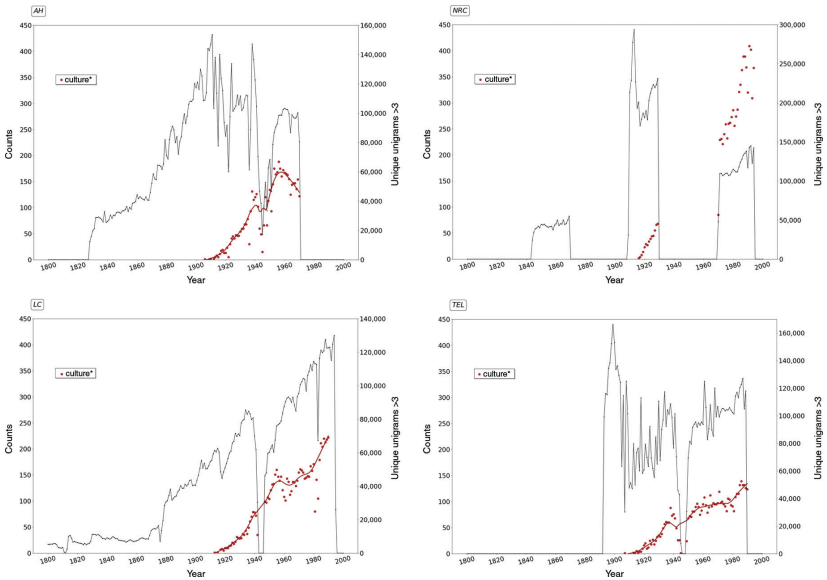

Figure 10. Bigram productivity for culture* *as adjective, in articles in* AH*, *NRC*, *LC and* TEL* between 1800 and 1990; only words of more than three characters have been taken into account.*

- P4: rubber culture, java culture, colonial culture, dutch east indies (*insulinde*) culture, modern culture
- P5: dutch culture, european culture, japanese culture, a bit of (*broodje*) culture, american culture

Vectors. In the same way as outlined above, we examined the changing semantic context of 'civilization' and 'civilized' using vector space models based on articles that included the words *beschaving** or *beschaafd**. Again, we calculated the most similar terms for each model and selected those terms that appeared in more than one model. This results in a chronological overview of words most similar to 'civilization' (Fig. 11).

It is possible to discern three groups of meaning over the five periods. In the nineteenth century (P1 and P2), 'civilization' is mostly about enlightenment and prosperity. The terms 'people's development' and 'affluence' feature in both periods, pointing to progressive cultural and economic development. This sense of collective enlightenment is further exemplified by such words as 'people', 'mankind', and 'humanity' and by 'industriousness', 'kind-heartedness' and other words denoting virtue. Civilization, in other words, was something to be pursued by humanity as a whole, apart from being a concept tied to geographical space.[36]

In P2 the concept of civilization contributes a heightened sense of purity and superiority, made clear by epithets like 'western', 'race', 'intelligence' and 'barbarians'. These terms can be construed as belonging to a colonial frame of thought. This need not always be the case, however: words such as 'civilized' and 'polite' were applicable also to social class. In P3, this normative dimension remains dominant. Words such as 'riches' and 'science', which emerge after World War II, indicate a return to the idea of civilization-as-prosperity, although a sense of civilization as a set of globally shared values is also present (with 'freedom', 'democracy' and 'community'). P5 saw the introduction of civilization as a more technical and neutral historical concept. Words such as 'history', 'human' and 'world' show that the concept was now used to discuss the rise of civilizations as part of a global historical narrative.[37]

Modern Europe: Entanglements

In a sense, Europe is the most recent member of the MCE trinity. Accounts of Europe as a concept stem from the nineteenth century,
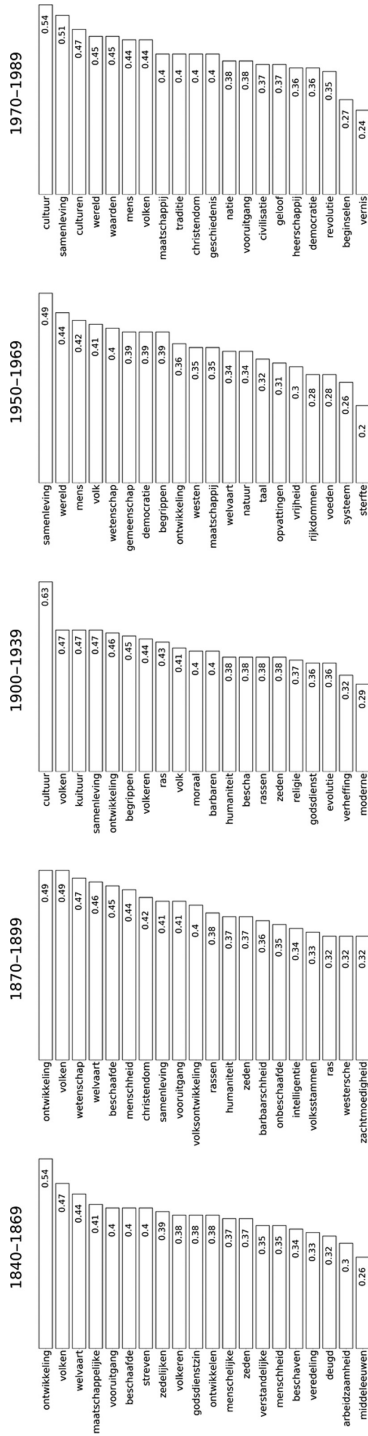
Figure 11. Words most similar to the adjective 'civilization' in periods P1–P5. The figure includes only the most similar words that are present in multiple models.

Table 6. Occurrences of bigrams for 'modern europe' and 'modern european(s)' per newspaper per period (absolute frequencies).

Newspaper	Bigram	P1: 1840–1869	P2: 1870–1899	P3: 1900–1939	P4: 1950–1969	P5: 1970–1990	Total
AH	modern(e) europa	0	9	21	2	1	33
LC	(modern europe)	0	4	4	6	7	21
NRC		1	0	17	0	30	48
TEL		0	0	28	6	16	50
Total		**1**	**13**	**70**	**14**	**54**	**152**
AH	modern(e) europe(e)	0	7	25	18	1	51
LC	s(ch)(e)	0	3	7	17	19	46
NRC	(modern european)	1	0	29	0	66	96
TEL		0	0	22	27	42	91
Total		**1**	**10**	**83**	**62**	**128**	**284**
AH	moderne europea(a)	0	0	0	5	0	5
NRC	n(en)	0	0	3	0	7	10
TEL	(modern europeans)	0	0	5	0	3	8
Total		**0**	**0**	**8**	**5**	**10**	**23**

receiving an enormous impulse after World War I and especially World War II. Europe was portrayed as a Christian continent, as the height of Enlightened modernity, as a diverse but coherent unity and as a conglomeration of nations undergoing a process of technocratic reconstruction.[38]

Bigrams and Collocations. To trace the conceptual entanglement claimed by the literature, we introduced 'Europe' into the equation, employing two approaches to examine the closeness of the relationship between 'the modern' on the one hand and Europe on the other. The first approach consisted simply of tracking the bigrams 'modern Europe' and 'modern European(s)' over time. The results are not spectacular (Table 6): even the number of absolute frequencies is extremely low.

TEL and *NRC* have the highest counts for 'modern europe', about fifty hits in all, although *NRC* attains this level in less than half the number of issues. The bigram occurs haphazardly over time in all newspapers. 'The Chinese Against Modern Europe', ran one typical headline.[39] The bigram 'modern european', as in 'modern European [military] strategy and equipment',[40] exhibits a similar pattern: the *NRC* scores disproportionately high and the bigram is spread out over time.

N-grams of higher value entail greater semantic specificity and could perhaps point to a stronger sense of entanglement. Given a cut-off point at a frequency of less than five hits (lower frequencies have been discarded), there is, however, not a single trigram, quadragram or pentagram that straightforwardly represents the MCE trinity. This the entire list of pentagrams containing 'modern', 'civilized' or 'civilization' in addition to 'Europe' in *AH*:[41]

– biggest and most modern in Europe (*grootste en modernste van europa*)
– one of the most modern of Europe (*een der modernste van europa*)
– to the most modern of Europe (*tot de modernste van europa*)
– of the most modern of Europe (*van de modernste van europa*)

– all civilized countries of Europe (*alle beschaafde landen van europa*)
– the civilized countries of Europe (*de beschaafde landen van europa*)

- the civilized peoples of Europe (*de beschaafde volken van europa*)
- most civilized countries of Europe (*meeste beschaafde landen van europa*)
- all civilized states of Europe (*alle beschaafde staten van europa*)
- of the civilized states of Europe (*der beschaafde staten van europa*)

- and the civilization of Europe (*en de beschaving van europa*)
- the civilization of Europe and (*de beschaving van europa en*)
- for the civilization of Europe (*voor de beschaving van europa*)

A broader approach to n-grams is afforded by collocations. They allow us to cast the net wider, making it possible to determine whether the close occurrence of words such as 'modern' or 'most modern' and 'Europe' is in any way compelling. Understanding collocations to refer to pairs of words that occur more often together than would be expected if the words had been randomly scattered across the newspaper articles, Europe and modernity do not seem very strongly related (Fig. 12).[42]
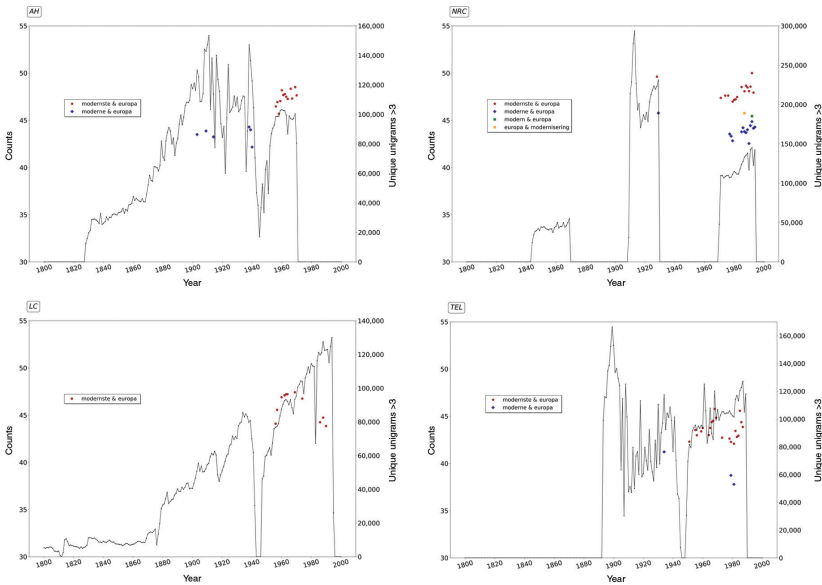
Figure 12. Collocates of modern* *and* europa*, *in articles in* AH*, NRC*, LC *and* TEL *between 1800 and 1990; only words of more than three characters have been taken into account.*

The exception is the superlative 'most modern', which occurs almost exclusively in the post-war years. However, 'most modern' does not relate specifically to Europe, but more often than not to modern things (technologies, buildings, factories) within Europe. The following examples illustrate this:[43]

- ... that this Dutch company will be the *most modern* brewery in Europe when the new building is ready (*NRC* 1928)
- The Apollo Variety Theatre, which used to be a famous place of entertainment ... is now being converted to a cinema, which is to be the *most modern* in the whole of Central Europe (*NRC* 1929)
- If Europe would then no longer be capable of making the *most modern* chips, the whole European electronics industry is doomed to disappear (*NRC* 1989)

There is, then, very little to indicate that an explicitly normative euro-centrism was hardwired into Dutch language use. It is therefore time to look at more implicit manifestations of the MCE trinity.

Vectors. We have already seen what simple diachronic representations of word embeddings can tell us about the semantic context of words, and consequently about the interrelatedness of the elements in the MCE trinity. Yet a robust relationship between Europe and modernity is hard to find. Searches for terms most similar to 'Europe' or 'European' mostly yield words that can be categorized as spatial, such as 'continent' or 'mainland', and socio-political, such as 'peoples' and 'tribes'.

Because the constellation of most similar terms can be interpreted in terms of networks and clusters, we employed network analysis to take a closer look at the entanglement of Europe and modernity. We hypothesized that entanglement, if it existed, should be visible in an overlap of words that are most similar to 'modern' and 'european'. These words, along with the words most similar to them, can be visualized as a network (Fig. 13).

The conceptual entanglement of Europe and modernity evidently only emerges in P3, where 'western' forms the direct link between 'european' and 'modern'. The entanglement continues in P4 and P5, where 'new' also serves as a link.

Thus far the network-based analysis of word embeddings is based on the most similar terms to the three input words of the MCE trinity. This results in a fairly simple network. Especially with key concepts
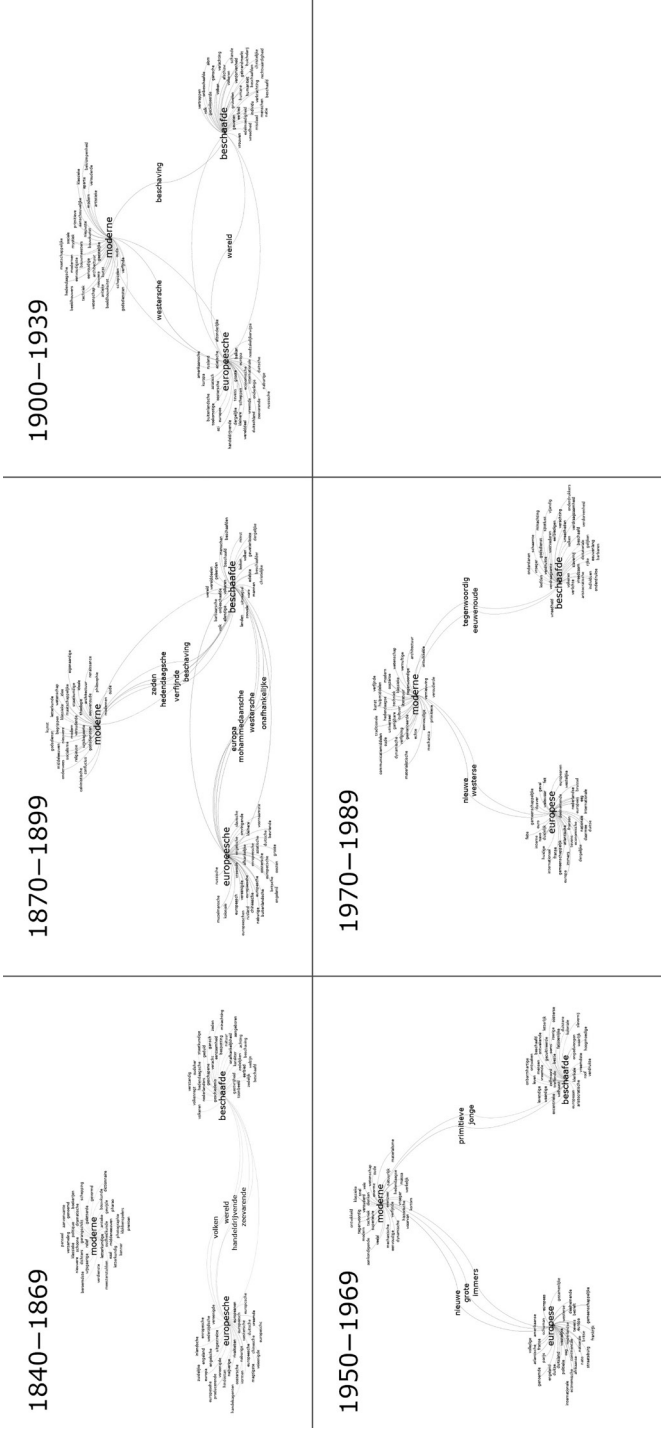
Figure 13. A network overview of the thirty words most similar to the adjectives 'modern', 'civilized' and 'european' in P1–P5. The network reveals (limited) entanglement of words similar to multiple concepts.

1840–1869

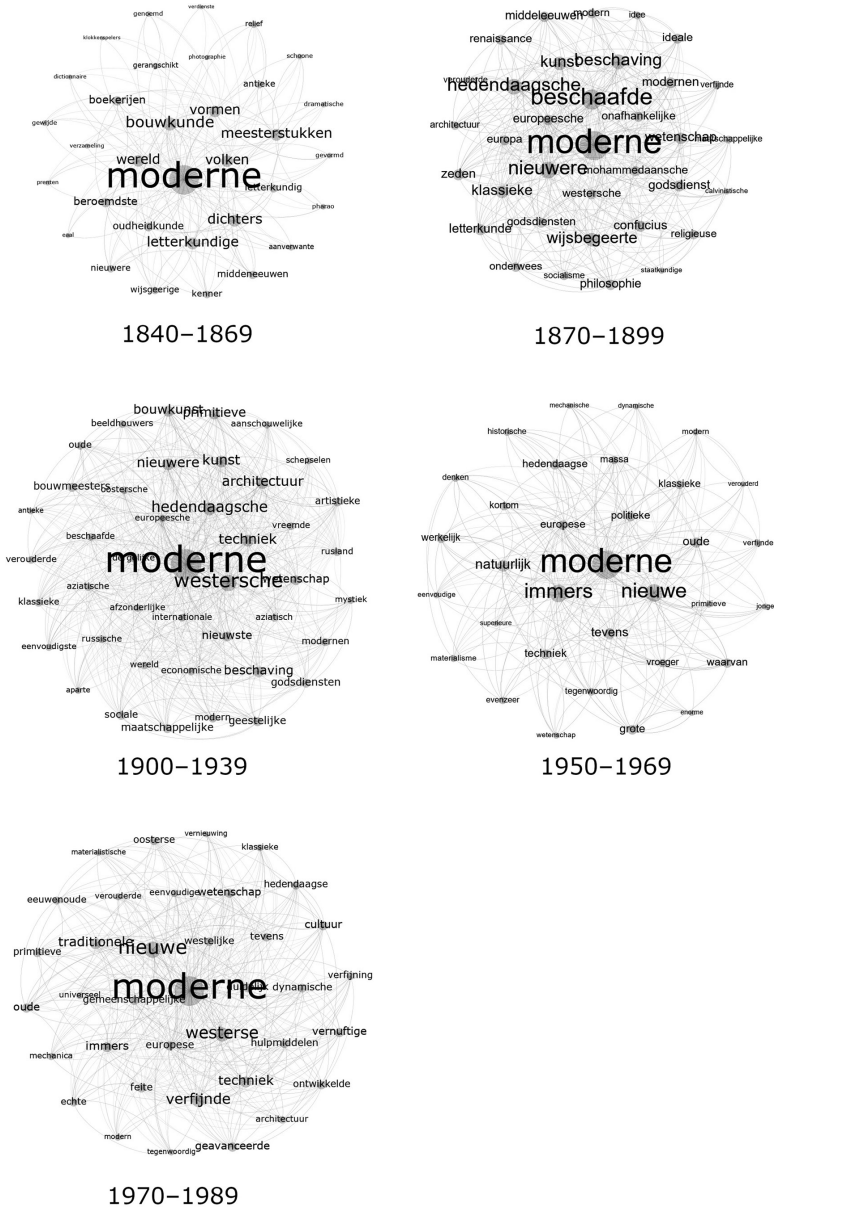1870–1899

1900–1939

1950–1969

1970–1989

Figure 14. A network overview of conceptual entanglement. By aggregating the thirty words most similar to the thirty words most similar to 'civilization', 'civilized', 'modern', 'european' and 'europe' a broader semantic field can be mapped, including the concepts that are semantically 'central' in the field. Only words with a connectivity of >30 have been included.

such as civilization and modernity, the analysis would benefit from more nodes (that is, words) in the network. For this reason, we included a second, deeper level, by calculating the most similar terms of each of the most similar terms that followed from modern, civilized and european.[44] The resulting 13,500 words allow us to investigate the entanglement on the basis of the network 'degree' (of connectivity), in addition to the semantic focus of the MCE trinity (Fig. 14).

As becomes clear from the network, the entanglement between Europe and modernity is, again, limited. While 'european' appears in four out of five periods, it does not show up as a particularly important node in the network. In P1 'european' is not even included in the network (meaning that it has a degree score of less than 30) and in P2–P4 the adjective has relatively low degree scores. This tells us that, if the conceptual MCE trinity existed, the concept of Europe was only a minor part of it. Europe was (closely) connected to the concept of modernity, but so were other concepts. Of course, one might argue that science and art, for example, were European phenomena, implying a more implicit presence of Europe in the trinity. We argue, however, that the word embeddings and their representation as networks largely take this implicitness into account. If 'modern' were, in fact, a 'hidden synonym' for 'Europe', both words would have been more closely entangled in the models because they would have been used in similar contexts by definition. In fact, the high scores of 'western' in several periods shows that even if we think of Europe as a spatial category implicit in modernity, it is still less significant than similar spatial (or socio-political) categories.

Civilized Europe: Entanglements

N-grams and Collocations. As we saw above, the bigram 'european civilization' is one of the more common ones in the whole corpus. It exhibits a specific pattern: the term occurred more frequently before World War II and clearly peaks between 1939 and 1941, when, apparently, fewer newspapers had more to say about European civilization than ever before. 'Western civilization' has a similar pattern, and so do 'European' and 'Western culture'. The latter is almost exclusively a twentieth-century affair, peaking in and around the World War II. In all cases, *NRC* scores the higher relative frequencies (Fig. 15 and Fig. 16).
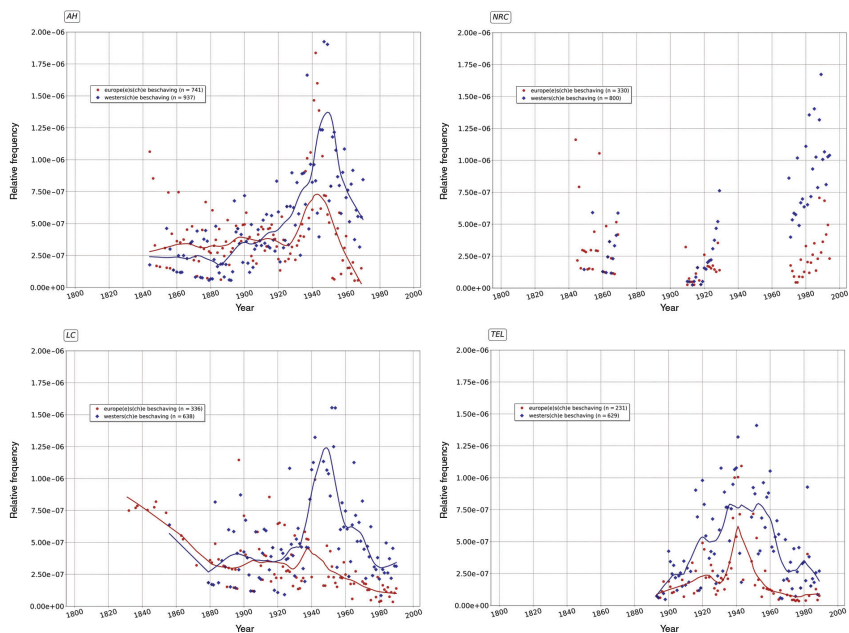
Figure 15. Bigrams of europe(e)s(ch)e beschaving *and* westers(ch)e beschaving, *in articles in* AH, NRC, LC *and* TEL *between 1800 and 1990.*

Collocations throw a slightly different light on the entanglement of MCE as a conceptual trinity. It is evident from the graph displaying the collocations of the pairs 'European' and 'civilization' on the one hand, and 'Western' and 'civilization' on the other, that the occurrence of both pairs significantly stand out from the total dispersion of words in documents containing the word *beschaving* (Fig. 17). The PMI index for post-war collocations appears to be slightly lower than those from before the war.

The same graph for 'culture' as one of the paired words (Fig. 18) shows a different pattern, however.

The pairs occur together in very different ways across the newspapers, appearing prominently in both *AH* and *NRC*. In *LC*, 'European' and 'culture' is a significant combination only in the second half of the nineteenth century, while 'Western' and 'culture' is exclusively a twentieth-century phenomenon. For *TEL*, on the other hand, 'European' and 'culture' are relevant only in the post-war period. The more common collocates of 'european' in articles containing *beschaving* (excluding
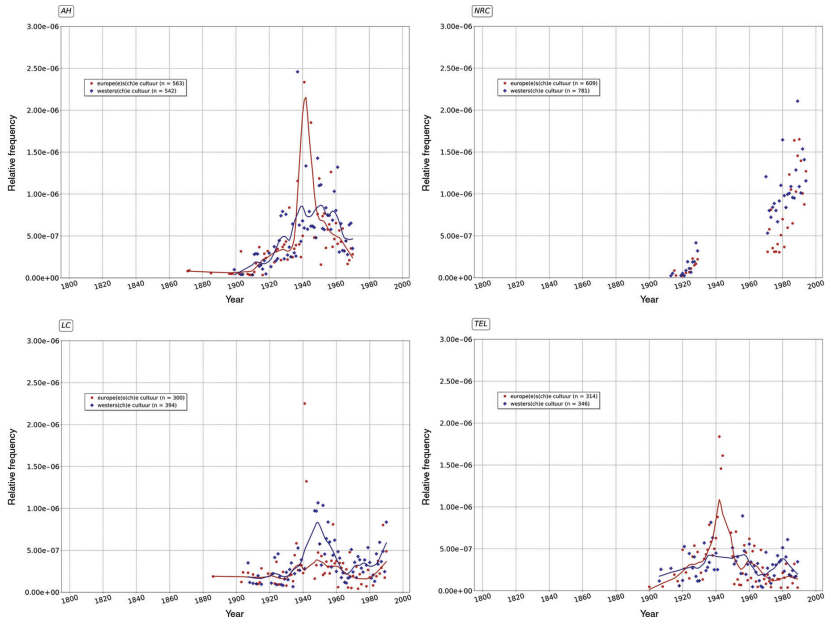
Figure 16. Bigrams of europe(e)s(ch)e cultuur *and* westers(ch)e cultuur, *in articles in* AH, NRC, LC *and* TEL *between 1800 and 1990.*
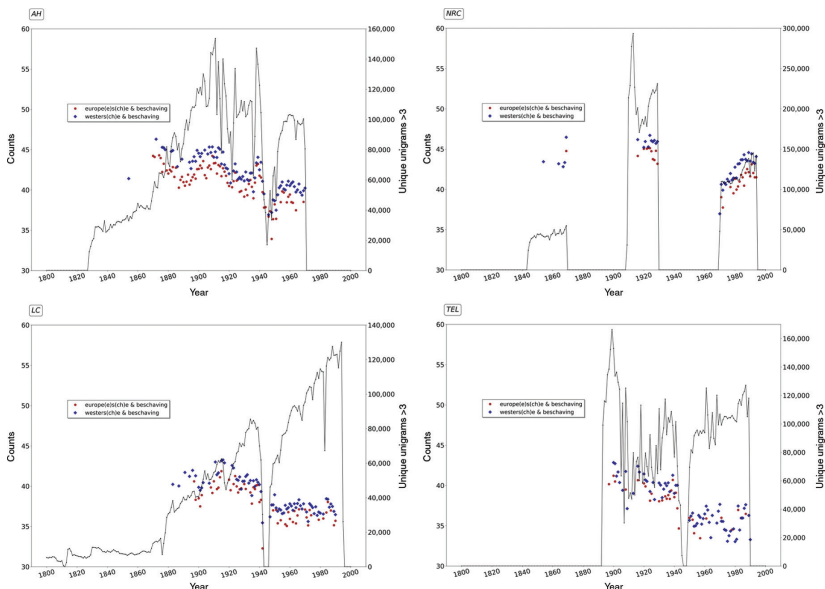

Figure 17. Collocates of europe(e)s(ch)e & beschaving *and* westers(ch)e & beschaving, *in articles in* AH, NRC, LC *and* TEL *between 1800 and 1990.*
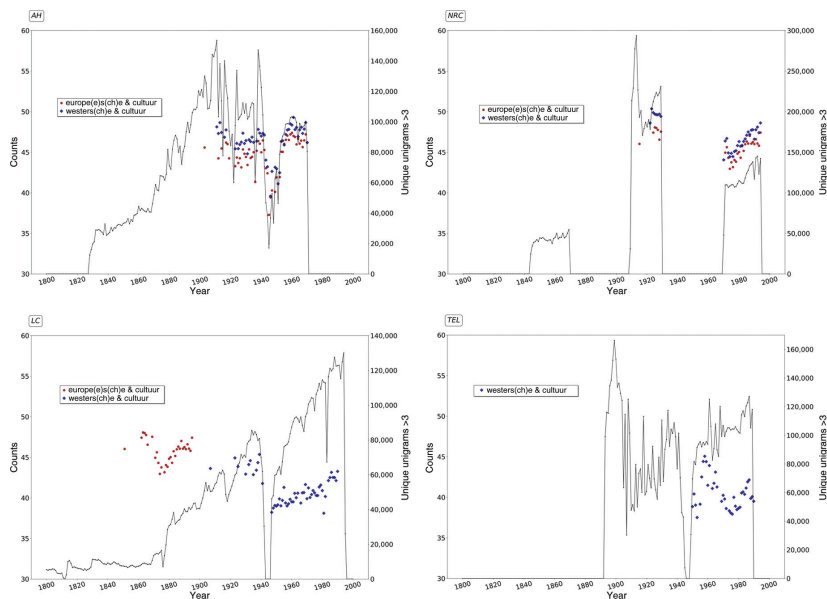
Figure 18. Collocates of europe(e)s(ch)e & cultuur *and* westers(ch)e & cultuur*, in articles in* AH*,* NRC*,* LC *and* TEL *between 1800 and 1990.*

the word *beschaving* itself) are mostly concerned with political topics: in *AH*, for instance, 'powers', 'countries', 'states' and 'peoples' emerge in P2, to which list are added 'civil servants', 'society' and 'education' in P3, and 'unity', 'spirit' and 'integration' in P4. The other newspapers show essentially the same pattern, expanding the list with words like 'community' and 'parliament'. Collocations of 'european' in articles containing *cultuur* (excluding the word *cultuur* itself) in *NRC* are nouns like 'administration' (P1), 'countries', 'personnel', 'businesses', 'states', 'education' (P3), and 'community', 'commission', 'parliament', 'art', 'cooperation', 'history', 'integration' and 'concert stage' (P5). The collocations for culture in *AH* are no different (*LC* and *TEL* have no collocations at all).

Vectors. Compared to 'modern', 'civilized' seems to be more strongly connected to the concept of Europe. Fig. 14 shows how nodes connect the two concepts in vector space in the first three periods, and to a certain extent also in the fourth period (through 'a european'). It is striking how the link between the two concepts develops between

P1 and P2. This change follows the distinction between civilization as enlightenment cum prosperity, and civilization as superiority cum purity. Between 1840 and 1869 Europe and civilization were connected through their spatial connotations. Both Europe and civilization refer to the 'sea-faring' and 'trading peoples' of Europe. The large number of positive elements in the civilization cluster ('independence', 'patience', 'morals') can be interpreted as a sign that civilization primarily involved European self-representation. In P2, this identity-forming capacity expands and, moreover, the concept is used in contradistinction to 'non-civilized'. Words most similar to 'civilized' now include negative terms such as 'ruthless', 'cruelty' and 'barbarian'. This identification of the lack of civilization with the non-European clearly became spatialized, as is evident from the occurrence of spatial categories ('continents', 'peoples', 'countries') that are close to 'civilized'. From P1 onwards, civilization as a set of values shared by humanity is more common and it seems that the relation between Europe and civilization in the twentieth century should be understood in this way.

As noted previously, 'western' always precedes 'european'. In fact, from their proximity in vector space it becomes clear that from P1 onwards 'western' was more related to both modern and civilized. Cosine distances between 'western-civilized' and 'western-modern' become more relevant than those between 'european-civilized' and 'european-modern' from P3 onwards, in models trained on articles based on *europa* – where one would, in fact, have expected a clear bias towards europe.

Conclusion and Discussion

Where has this digital history approach to the history of the concepts 'modernity', 'civilization' and 'Europe' brought us? In what follows we discuss three outcomes. Firstly, as expected, our analysis generally supports the findings of more traditional approaches to conceptual history. Secondly, we have been able quite precisely to show how the concepts developed, in respect to their timing, the quantities in which words that denoted them occurred, and the differences in language use across the newspapers. Thirdly, we have identified a 'Eurocentric fallacy'. Although it is obviously possible to flesh out this topic in much greater detail still, it transpires that the mere equation of the members of the MCE trinity is not as straightforward as the literature suggests.

The congruence between our digital approach and previous findings can be illustrated by using modernity as an example. For this we have the classic accounts by scholars like Reinhart Koselleck and Hans Ulrich Gumbrecht. As we saw, in P1 (but in fact since at least 1800), *modern** was predominantly used to denote the contemporary as opposed to the historical or old. Bigrams such as 'modern exterior' or 'painted in modern fashion' indicate an understanding of the modern as a condition of newness. These novelties share a common semantic property, in that their qualifying adjective 'modern' is predominantly used in a practical and material sense. Throughout the nineteenth and twentieth centuries, 'modern' frequently co-occurs with concrete items such as 'kitchen', 'house' and 'furniture'. Modernity, it appears, was largely sustained in public discourse through consumer goods and technology.[45]

The usage of 'modern' expands after the mid-nineteenth century, but it also changes. Not only did the adjective become attached to an increasing number of nouns, from the 1860s onwards a process of abstraction also began. 'Modern' began to qualify immaterial phenomena and ideas. This process of abstraction can be read as an instance of what Koselleck described as 'condensation'. On its way towards becoming a 'collective singular', the modern became attached to large spatial and social categories such as the state, civilization and the world. For instance, the relatively short period between 1860 and 1880 saw the introduction of bigrams such as 'modern civilization', 'modern freedom', 'modern spirit' and 'modern world'.

The concept of modernity also gained a more normative connotation. The modern was no longer simply understood as the 'present' as opposed to what went before, but also as the 'new' (and therefore 'better') in contrast to the 'old' (and therefore 'outdated'). The bigrams so plenteously introduced in P3 ranged from 'modern women' to 'modern radio' and 'modern neighbourhood'. Consistent with the spread of technology, consumer goods and scientific innovation, the vocabulary of modernity grew extensively. More and more aspects of social life were enunciated as 'modern'; there seems, in fact, to have been no limit to what could be considered as such. Key to understanding this stage in the conceptual evolution of modernity is Koselleck's hypothesis of conceptual democratization, even though P3 is far past the so-called *Sattelzeit*. In the form of cars, fashion, voting rights and popular culture, the fruits of modernity became accessible to more and more people, and language reflects this democratization of the modern. The

period in question also saw the identification of the concept of modernity itself as 'modernization', 'modernism', and compounds such as 'ultramodern' and 'hypermodern' began to be applied to many aspects of society, politics and the economy.

The post-war decades witnessed a reassessment of the more abstract nouns introduced a century earlier. Especially between 1950 and 1960 bigrams such as 'modern civilization', 'modern state' and 'modern times' occur frequently. There was a gradual rise in productivity, with new bigrams directly reflecting the reconstruction of a *modern* society. Bigrams such as 'modern american', 'modern industrialized' and 'modern armaments' indicate how the adjective was specifically geared towards describing economic reconstruction, consumer goods and international politics (in particular the Cold War). In general, usage of the word 'modern' points towards the future-oriented outlook inherent in the term throughout the nineteenth and twentieth centuries. Lacking, however, are several important claims with regard to the conceptual history of modernity, such as Charles Baudelaire's take on modern life as fleeting and ephemeral, the ideological colouring of the different futures envisioned by different historical actors, and the extreme acceleration of time around 1900 characteristic of avant-garde modernity.[46] These absences are important to keep in mind, since they demonstrate both the weakness and the strength of newspapers as a source for conceptual history. If big data dilutes deeper philosophical meanings to the point of oblivion, it also reflects the more superficial but popular understandings that are society's greatest common denominator.

Secondly, as the representation of our findings amply demonstrates, a digital history approach offers a precise indication of how word usage developed over a longer period of time. This applies specifically to the use of n-grams and collocations, which can be examined on the level of years or even months. Word embeddings require much more data in order to attain a plausible degree of precision. On the other hand, vector space distances, PMI measures and n-gram frequencies can be quantified. Despite the error margins caused primarily by imperfect optical character recognition (OCR), we have been able to offer a fair idea of the numbers of unigrams and bigrams involved, and perhaps more importantly, of the intellectual productivity and rhetorical creativity the new concepts engender.

This article has also brought to light distinctions across the periodical media. One of the interesting findings concerns the relatively

high presence of MCE elements in *NRC* and to a lesser degree *AH*. In regard to bigram frequencies, bigram productivity as well as collocations the scores of *AH* and *NRC* stand out. We assume that these scores result from greater interest in 'civilization' as a concept popular among most conservatives since the nineteenth century. Both *AH* and *NRC* were newspapers that tended towards the right. This ideological position was especially prominent in the 1970s and 1980s, a period known for its left-wing/right-wing polarization. At the same time, as a self-styled quality newspaper, *NRC* (as *NRC Handelsblad*) catered predominantly to the intellectual elite. In this sense, its productivity in generating bigrams might well be the result of using a larger vocabulary in combination with a need to express rhetorical prowess. This may explain the difference with *TEL*, another right-wing newspaper, but a more popular ('tabloid' style) one that served a different audience.

A third outcome of our examination of the MCE trinity is that even with the inclusion of more or less right-wing newspapers in the dataset, the conceptual entanglement within the purported MCE trinity is highly restricted. As we have seen, connections are limited to only a handful of words. The scarcity of entanglements is confirmed by the number of words that are *not* in the network. Only 0.98%–1.33% of the 13,500 output words prove to be connected to more than thirty other words. Also, in P2 (the period with the highest degree of entanglement) the network 'density' (the number of actual connections divided by the number of possible connections) amounts to a meagre 0.539, even after selecting only words with a degree score higher than thirty. Moreover, within the MCE trinity practically no connections exist between all three elements. In P1 there are connections between 'civilized'/'european', in P2 between european/'civilized' and 'modern'/'civilized', in P3 between all three elements but only through a couple of words, and in P4 and P5 between 'european'/'modern' and 'modern'/'civilized'. The result does not convincingly prove the persistent presence of an MCE trinity.

This, then, answers the question we posed. Instead of an extensive and constant intertwinement of the MCE elements, our research shows intermittent and alternating connections. The MCE trinity was never an equally balanced triangle (Fig. 15); in fact, it barely qualifies as a triangle at all. While modernity, civilization and Europe have always shared a discursive border, as the literature claims, the border itself is blurred by a semantic no-man's-land of substantial proportions.

Europe, civilization and modernity certainly shared similar pasts, but they followed different semantic trajectories and gave rise to different semantic configurations. There is no reason to suppose that different national contexts will give rise to different results, but that, of course, remains to be proven. At best, our results suggest that if differences are found, they would be differences between newspapers with dissimilar identities. Our research suggests that the 'West' rather than 'Europe' may have been of greater significance in connection with both modernity and civilization; but this, too, needs to be demonstrated.[47]

The conclusion that the MCE trinity was perhaps not perceived as a trinity in our sources follows from the kind of research we have done, focused as it is on frequencies and syntagmatic relations. Explaining why entanglement was so limited requires a move beyond a mere token and frequency-based analysis. Collocations and ngrams reveal a part of the referential aspect of our concepts, but contexts on the level of articles or even paragraphs remain hidden in distant reading methods. Another potentially interesting aspect is the rhetoric of modernity, civilization and Europe. Tracing the argumentative structures in which members of the trinity feature would greatly enlighten our understanding of their usage and their interrelationships.

Also, our conclusions follow from the kind of source material we have opted to use. Nineteenth and twentieth-century newspapers are a far cry from the intellectually dense reflections one finds in learned treatises and books. Yet the conceptual-cognitive realm of intellectual thought is necessarily interlinked with the common prose of periodical media. It has been debated practically since their inception whether newspapers represent public opinion, and no definitive answer has been forthcoming. No-one, however, will dispute that periodical media represent public opinion to a greater degree than do studies of intellectual depth. Yet scholars, hampered by the 'streetlight effect', have tended to seek answers where the light is, leading them to conclusions that we have identified as a Eurocentric fallacy.

Appendix

Data quality. The corpora we have used come with imperfections, in particular as regards their quality. Problems with digitized periodicals

come on three levels: spelling variation, spelling change, OCR distortion and segmentation.[48] Spelling variation is not a major issue for this period. Nor is spelling change, as the only important change occurred between 1947 and 1955, which we have taken into account in our periodization.[49] OCR distortion is a much larger problem. For this reason, we have not worked with advertisements (although it is impossible to divest the dataset completely of all non-articles, as the automatic segmentation of newspapers used in constructing the dataset too is imperfect). However, none of the words we have worked with is problematic per se; for example, they contain only the standard alphabetic characters (that is, no ligatures or other special characters). As far as we have been able to ascertain on the basis of lists of unigrams per year, OCR problems related to the keywords we have used are more or less evenly distributed over time, so that errors are balanced and comparisons possible.

We have partially cleaned the dataset by rejecting tokens (a series of characters, in this case 'words') that occur infrequently but account for a large number of errors, using relatively low cut-off points at a frequency of 4. A rough calculation shows that the error margin in determining the unigram 'europe' is surprisingly small, suggesting a margin of 2%.[50] An error margin of 5–10% seems safer, however, especially when tracing change over time. The transition from the prewar to the postwar period, for example, saw an increase in the OCR quality, which in some cases was rather substantial. The dataset with the best OCR quality of newspapers in our dataset is, incidentally, *LC*.[51] A more pressing problem for future research on many datasets, including this one, is article segmentation. Currently it is impossible to differentiate clearly between different types of articles, although probabilistic methods have been developed to categorize articles into genres such as 'editorial', 'news report' or 'opinion'.[52]

Vector space models. Vector space models are most reliable when 'fed' with large quantities of data. Because the number of words per year and per decade was too small to train separate models, we divided our time frame into five blocs (P1–P5). We subsequently took the size of the period with the least data (P1, about 500 MB of articles) as limit for the other period.[53] The articles used were articles that mentioned keywords related to 'europe', 'civilization' and 'modernity'.[54] Ultimately, we trained 5 (number of periods) × 5 (number of keywords)

= 25 different models. Diachronically comparing the models requires so-called 'alignment'; models can be composed of similar words, but the structure and shape of the model is always unique. For this reason, we used a method of harmonizing/aligning the models suggested by Hamilton et al.[55] After alignment, the models are fit for comparison, and diachronic semantic change can be analyzed.[56]

Our method of selecting data, training models and aligning them is not unproblematic. First, subsetting the data based on keywords results in a topicality bias; the word civilization is likely to be more 'european' in a model trained on articles that mention 'europe'. Still, the lack of entanglement visible in our results indicates that this problem does not significantly change our results. A more pressing issue is the fact that searching for 'most similar terms' to word x in a vector model trained on articles mentioning word x simply yielded less meaningful output terms. Because of the high frequency of word x as the input word, search operations of most similar terms will return all OCR errors of the input word.[57] To counter this problem, as well as the topicality bias, we retrieved the most similar words to our input terms, but selected only the ones that appear in multiple models. For example, the most similar terms to 'civilization' in the second period were calculated by aggregating the (twenty) most similar terms to 'civilization' in the five models of period, after which we selected only the duplicate words and calculated their average cosine similarity. Lastly, the method used to align the models uses the vocabulary of the first period to align the models in subsequent periods. This prunes new words (not present in P1, but only in later periods). This method has its limitations, however; other methods have been proposed.[58] However, the most similar words for 'modern', 'civilization/civilized' and 'europe/european' in the *untrained* models in P2–P4 show that terms semantically similar to these input terms exist already in P1.

Notes

1   We would like to express our gratitude to Simon Hengchen, Pim Huijnen, Ed Jonker, Lorella Viola, Leon Wessels and the anonymous reviewers for their comments on previous versions of this article.

2   The literature on the subject is very large. Several recent articles in a recent edition of *History* summarize the *status quaestionis*: Gavin Murray-Miller,

'Civilization, Modernity and Europe: The Making and Unmaking of a Conceptual Unity', *History* 103 (2018), 418–33. Very similar points are made by Matthew D'Aura and Kan Vermeiren, 'Narrating Europe: (Re) Thinking Europe and its Many Pasts', in: *History* 103 (2018), 385–400 and Rolf Petri, 'Meanings of Europe and Meaning in History', in: *History* 103 (2018), 401–17. A range of well-known studies have tried to salvage the trinity from postmodern criticism, accepting Europe's 'provincialized' status while highlighting its distinctive civilizational character. A classic statement is Shmuel N. Eisenstadt, 'The Civilizational Dimension of Modernity: Modernity as a Distinct Civilization', in: *International Sociology* 16 (2010), 320–40; for an overview of this discussion, see Gerard Delanty, *Formations of European Modernity*: *A Historical and Political Sociology of Europe* (2nd ed.; London, 2018).

3   Murray-Miller, 'Civilization, Modernity and Europe', 424.

4   Bo Stråth and Peter Wagner, *European Modernity*: *A Global Approach* (London etc, 2017), 5.

5   Ernst Müller and Falko Schmieder, *Begriffsgeschichte und historische Semantik. Ein kritisches Kompendium* (Berlin, 2016) 783–801, on 'Begriffsgeschichte im digitalen Zeitalter'; Silke Schwandt, 'Digitale Methoden für die Historische Semantik. Auf den Spuren von Begriffen in digitalen Korpora', *Geschichte und Gesellschaft* 44:1 (2018), 107–34. Groups active in the field include http://www.crassh.cam.ac.uk/ programmes/the-concept-lab-cambridge-centre-for-digital-knowledge and https://comhis.github.io/ (accessed 13 August 2019).

6   Rolf Reichardt, Eberhard Schmitt and Brigitte Schlieben-Lange eds., *Handbuch politisch-sozialer Grundbegriffe in Frankreich*, *1680–1820. Heft 1–2* (Munich, 1985) 22–40.

7   Tom Kenter, Melvin Wevers, Pim Huijnen and Maarten de Rijke, 'Ad Hoc Monitoring of Vocabulary Shifts over Time', in: *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management* (*CIKM '15*) (New York 2015) 1191–1200; Gabriel Recchia, Ewan Jones, Paul Nulty, John Regan and Peter de Bolla, 'Tracing Shifting Conceptual Vocabularies Through Time', *European Knowledge Acquisition Workshop* (2016) 19–28.

8   https://www.delpher.nl; the data was extracted through the APIs made available by the Koninklijke Bibliotheek (https://www.kb.nl/en/resources-research-guides/data-services-apis, accessed 13 August 2019).

9   An informed overview of Dutch newspapers is Huub Wijfjes, *Journalistiek in Nederland 1850–2000. Beroep*, *cultuur en organisatie* (Amsterdam, 2004).

10 It will be evident from Table 1 that articles were of longer length in P3 than in the other periods. Because of this discrepancy we have opted to normalize the data (i.e. determine relative frequencies) by using the number of tokens rather than the number of articles. The average ratio of the number of tokens to the number of articles per period is: 38.37 (P1), 34.98 (P2), 75.25 (P3) 34.89 (P4), 28.14 (P5). We have used relative rather than absolute frequencies wherever pertinent.

11 Melvin Wevers, 'Consuming America. A Data-Driven Analysis of the United States as a Reference Culture in Dutch Public Discourse on Consumer Goods, 1890–1990', unpublished PhD dissertation, Utrecht University, 2017 (https://dspace.library.uu.nl/handle/1874/355070), 77.

12 Compare Jim Berryman, 'Civilisation: A Concept and its Uses in Australian Public Discourse', *Australian Journal of Politics and History* 61/4 (2015), 591–605: 'Although the language of civilisation is not confined to centre-right political discourse, it has been most vocal among conservative-leaning commentators'.

13 Reinhart Koselleck, 'Begriffsgeschichte and Social History', *Economy and Society* 11:4 (1982) 409–27.

14 These are *europe(e)s(ch)(e)* ('european'), *beschaafd(e)* ('civilized', supplemented with *culture(e)l(e)* for 'cultural') and *modern(e)* ('modern').

15 For the calculation of PMI, see e.g. Daniel Jurafsky and James H. Martin, *An Introduction to Natural Language Processing*, *Computational Linguistics*, *and Speech Recognition* (third edition, draft: https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf, accessed 13 August 2019), Chapter 6.

16 See e.g. Andrey Kutuzov et al., 'Diachronic word embeddings and semantic shifts: a survey' in [*cs.CL*], (2018) https://arxiv.org/abs/1301.3781.

17 For a general conceptual history of 'modernity', H.U. Gumbrecht, 'A Conceptual History of the Word Modern', idem., *Making Sense in Life and Literature* (Minneapolis, 1992) 79–110; M. Calinescu, *Five Faces of Modernity*: *Modernism*, *Avant-garde*, *Decadence*, *Kitsch*, *Postmodernism* (Durham, 1987) 13–86.

18 Gumbrecht, *Making Sense*, 96.

19 *AH* 15-11-1878: 'zes jaren lang waren de mannen der reactie aan 't werk, poogden Frankrijk te vormen volgens hun eng begrip; ze wilden de moderne beschaving dooden'; *LC* 13-09-1880: 'De moderne beschaving doet dus, ook op het gebied van het volkenregt, de schoonste vruchten rijpen'; *AH* 19-07-1895: 'Welke onbeschaafde, woeste natie kunnen wij brengen onder den verlichtenden invloed der moderne beschaving, door

haar te bestralen met de zoeklichten van een oorlogskader?'; *LC* 05-10-1904: '... dat in het belang van den wereldvrede en de moderne beschaving het uiterste oogenblik is gekomen om dezen last van internationale verantwoordelijkheid op zich te nemen'.

20 These are *modern*(*e*), *moderner*(*e*), *modernst*(*e*).

21 *LC* 27-2-1818: 'de professionele boekerij van Buonaparte. Het zijn meest alle moderne boeken, en drukken van de kostbaarste uitvoering'; *AH* 25-10-1841: 'als door eene beschaafde opleiding in moderne talen en andere wetenschappelijke kundigheden, tot nuttige leden der maatschappij trachten te vormen'; *NRC* 14-12-1849: 'ons te verzetten tegen de suprematie, welke door den staat, naar moderne begrippen, dikwerf met een loden schepter over alle gezindheden uitgeoefend wordt'; *NRC* 9-8-1855: 'behalve onderwijs in de wis- en zeevaartkunde, ook onderwijs zal worden gegeven in de oude en moderne talen'; *NRC* 28-12-1868: 'alsof de moderne staat reeds onherroepelijk over den antieke had gezegevierd'; *AH* 14-12-1868: 'dat onze gansche moderne staat een gewrocht is van ongeloof en revolutie'; *AH* 1-12-1904: 'men ziet er uit, dat wij gelukkig niet op alle punten van modern verkeer achterlijk zijn'; *AH* 22-2-1910: 'een behoud van dat oude te vereenigen met de eischen van 'n modern verkeer'.

22 *LC* 28-06-1907: 'De typische groenten-stalletjes zijn dus verdwenen, maar er voor in de plaats gekomen is de drukte van modern verkeer in 't hartje van de stad.'

23 The distance in vector space between two words is their 'cosine similarity'. The higher the figure, the stronger the similarity. Tomas Mikolov et al., 'Efficient estimation of word representations in vector space', *Proceedings of ICLR Workshops Track* (2013) arxiv.org/abs/1301.3781, 1–11, at 5.

24 The Dutch words mentioned are respectively *nieuwe*(*re*), *hedendaags*(*ch*)*e*, *klassieke*, *oudheid*, *roman*, *scho*(*one*), *schilder*, *meesterstukken*.

25 The Dutch words mentioned are respectively *renaissance*, *gothische*, *middeleeuwen*, *eeuwenoude*, *verouderd*, *kerk*, *materialistische*, *architectuur*, *primitieve*, *geestelijk*, *verfijnd*, *techniek*, *geavanceerd*, *natuurwetenschap*.

26 'Onze foto's geven op sprekende wijze het verschil weer tusschen modern en Oostersch Bangkok' (*AH*, 10-12-1931).

27 See: Reinhart Koselleck, 'Zur historisch-politischen Semantik asymmetrischer Gegenbegriffe', idem, *Vergangene Zukunft: Zur Semantik geschichtlicher Zeiten* (Frankfurt a. M., 1989) 211–59. See also Helmut Hühn, 'Die Entgegensetzung von "Osten" und "Westen", "Orient" und

"Okzident" als begriffsgeschichtliche Herausforderung', Ernst Müller (ed.), *Begriffsgeschichte im Umbruch*? (*Archiv für Begriffsgeschichte*, *Sonderheft 4*) (Hamburg, 2005), 59–67.

28 Riccardo Bavaj and Martina Steber eds., *Germany and 'The West'*: *The History of a Modern Concept* (New York, 2015) 8.

29 Joris van Eijnatten, 'Protestantse schrijvers over beschaving en cultuur', in P. den Boer (ed.), in *Beschaving*: *Een geschiedenis van de begrippen hoofsheid*, *heusheid*, *beschaving en cultuur* (Amsterdam, 2001) 255–85; Reinhart Koselleck, 'Neuzeit: Zur Semantik moderner Bewegungsbegriffe', in idem, *Vergangene Zukunft*, 300–48.

30 R.A.M. Aerts and W.E. Krul, 'Van hoge beschaving naar brede cultuur, 1780–1940', Den Boer, *Beschaving*, 213–220. Note the passage on p. 232, where on the basis of a single reference (a work of history from 1865) rests the rather sweeping claim that in the nineteenth century, 'the higher classes [*standen*] of Christian, modern Europe *were* the norm for "civilization"' (original emphasis). This particular study is based on careful analysis; the problem, however, is that such findings tend to be reified into a European mind-set spanning decades if not centuries.

31 *NRC* 10-04-1924: 'Wie kan bewijzen, bewijzen, dat *onze beschaving* achteruit zal gaan als 't Grieksch wordt afgeschaft?'; *NRC* 20-12-1924: 'Hetzelfde voor *onze beschaving*: zij heeft met "*civilisatie*" maar weinig gemeens, en zou eigenlijk dichter bij de Fransch-Engelsche politesse staan'; *NRC* 13-09-1924: 'Hij herinnerde aan den invloed, die de Italiaansche cultuur, de Italiaansche kunst op *onze beschaving*, op onze kunstenaars hebben gehad'; *NRC* 18-6-1924: 'Waar zou het heen met *onze beschaving*, dit woord in zijn diepste beteekenis genomen, indien aan het winstbejag geen andere grenzen werden gesteld dan de gewone strafwet geeft?'

32 *AH* 11-04-1906: 'Het bestaan van een blanke beschaving in Zuid-Afrika is afhankelijk van de ontwikkeling der Kaffers in boeren-gemeenten met plaatselijk zelfbestuur'.

33 *LC* 6-12-1924: 'De planter in de tropen, die ver van blanke beschaving is verwijderd, kan den slag hooren van het hart van het rijk'.

34 *LC* 19-11-1914: 'Natuurlijk, daar is Japan nog, en is er een vruchtbaarder veld om bekoorlijke mythen te delven dan het zonnige land der gracielijke geisha's, voor het met de invasie der Europeesche cultuur werd gestraft?'; *AH* 10-7-1925: 'De Europeesche cultuur moet thans zijn laatste troef uitspelen. Zij is voor altijd verloren als de slachting opnieuw zou beginnen'.

35 *AH* 2-7-1899: 'Immers, sedert het principe van Europeesch onderwijs aan de inlanders is doorgevoerd (sedert 1823) heeft men weliswaar de Westersche cultuur voor hen opengesteld, maar tevens hun alle reden gegeven om gelijkstelling met de Europeanen te vragen'.

36 The Dutch words mentioned are respectively *volksontwikkeling*, *welvaart*, *volk*, *,menschheid*, *humaniteit*, *arbeidzaamheid*, *zachtmoedigheid*.

37 The Dutch words mentioned are respectively *westerche*, *ras*, *intelligentie*, *barbaren*, *beschaafd*, *rijkdommen*, *wetenschap*, *vrijheid*, *democratie*, *gemeenschap*, *geschiedenis*, *mens*, *wereld*.

38 For an overview of recent literature, see Joris van Eijnatten, 'Something about the Weather. Using Digital Methods to Mine Geographical Conceptions of Europe in Twentieth-Century Dutch Newspapers', in *BMGN. Low Countries Historical Review* 134/1 (2019), 28–61.

39 *LC* 29-6-1899: 'De Chinezen tegen modern Europa'.

40 *AH* 6-10-1895: 'moderne Europeesche krijgskunde en uitrusting'.

41 Conceivably the newspaper column structure could act as a restriction on the identifiability of pentagrams; however, the number of tri and quadragrams is likewise limited.

42 Collocates were generated from quadragrams on the basis of a narrow window of 4 (left and right).

43 *NRC* 9-06-1928: '... dat dit Nederlandsche bedrijf, als het nieuwe gebouw gereed is, de modernste brouwerij van Europa zal zijn'; *NRC* 9-02-1929: 'Het Apollo-Variété te Weenen, dat vroeger een befaamd uitgaansoord was ... wordt thans verbouwd tot een bioscoop, die de modernste van heel Midden Europa worden moet'; *NRC* 4-01-1989: 'Als Europa dan niet meer in staat is om de modernste chips te maken, is de hele Europese elektronica-industrie tot verdwijnen gedoemd'.

44 This resulted in a total of 3 (input words) × 30 (most similar terms) × 30 (most similar terms) × 5 (periods) = 13,500 words.

45 The observation corresponds with Gumbrecht's observation that newness derives from 'certain qualities that comprehend [something] as homogeneous in all [its] complexity'; Gumbrecht, *Making Sense*, 81.

46 Gumbrecht, *Making Sense*, 102–3.

47 If there is a strong connection between 'the West' and 'modernity' then 'America' would probably figure prominently as part of the West; but this still leaves 'civilization' unaccounted for.

48 Michael Piotrowski, 'Natural Language Processing for Historical Texts', *Synthesis Lectures on Human Language Technologies* 5:2 (2012), 1–5.

49 Among other things, the suffix 'sch' was dropped in many instances. In our case, the adjective *europeesch* for 'European' thus became *europees*.

50 An examination of all unigrams (unique words) in *AH* with a frequency higher than 4 shows that possible variants of *europa* amount to about 5,000 errors of erroneously conjoined words (such as *europaamerika*) in comparison to approx. 255,000 correct spellings. This comes down to an error margin of 2%. Not that the number of errors caused by the inclusion of wholly unrelated words (such as *n*europat*hologie*) is extremely small.

51 Wevers, 'Consuming America', 82–3.

52 See e.g. http://lab.kb.nl/tool/genre-classifier (accessed 24 June 2019).

53 Although the size of the models does not influence comparability after alignment we decided to limit the data because of the restrictions to our computing power. We 'topped' the data by taking randomly sampled articles.

54 Keywords: *europa*, *europe**, *beschaafd**, *beschaving*, *modern**.

55 William L. Hamilton, Jure Leskovec and Dan Jurafsky, 'Diachronic word embeddings reveal statistical laws of semantic change', arxiv.org/abs/1605.09096 (2016; accessed 13 August 2019). The corresponding code can be found on GitHub: https://github.com/williamleif/histwords (accessed 13 August 2019). Because of the relatively small datasets we used for the model, we chose a window size of ten and expanded the number of iterations to eight.

56 One problem particular to the alignment method used by Hamilton et al. is the issue of new words. One aspect of alignment is the intersection of the models by their shared vocabulary. This means that words that appear in the second period or later are dropped. Tests with unaligned models showed that in our case this was not a problem since the most similar terms to our keywords all existed in P1.

57 By way of example: *beschaade*, *beschaaf* and *beschaafdenlÃ⁻d*, errors which occur once each in *AH* in 1844, obviously have the same meaning as *beschaafd**, which occurs with a frequency of 190 in that year.

58 Yoon Kim, Yi-I Chiu, Kentaro Hanaki, Darshan Hegde and Slav Petrov, 'Temporal Analysis of Language Through Neural Language Models', *Proceedings of the ACL 2014 Workshop on Language Technologies and Computational Social Science* (2014) 61–5.

About the Authors

Joris van Eijnatten (PhD 1993) is professor of cultural history at Utrecht University, the Netherlands. He works on various interrelated subjects, including the history of ideas, religion, media and communication, employing source material from the eighteenth century to the present. His publications include books and articles on toleration, Dutch religious history, the poet Willem Bilderdijk, sermons, the justification of war, media history and conceptual history. He specializes in digital history approaches to newspapers, periodicals, parliamentary records and other digitized texts, with a focus on tracing changing concepts over time. E-mail: j.vaneijnatten@uu.nl

Ruben Ros (MA) studied at Utrecht University and has a pronounced interest in digital conceptual history. In his Masters' thesis he explored the conceptual evolution of 'foreignness' in Dutch newspapers during the long nineteenth century. He has worked as research assistant at Helsinki Computational History Group (COMHIS) and the Digital Humanities Lab of the KNAW in the Netherlands. E-mail: ruben@rubenros.nl