

Alexander, M. and Struan, A. (2022) “In barbarous times and in uncivilized countries”:
two centuries of the evolving uncivil in the Hansard Corpus. *International Journal of
Corpus Linguistics*, 27(4), pp. 480-505.



Copyright © 2022 John Benjamins Publishing Company. Reproduced under a
[Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

For the purpose of open access, the author(s) has applied a Creative Commons
Attribution license to any Accepted Manuscript version arising.

<https://eprints.gla.ac.uk/279035/>

Deposited on: 9 September 2022

“In barbarous times and in uncivilized countries”

Two centuries of the evolving uncivil in the Hansard Corpus

Marc Alexander and Andrew Struan

University of Glasgow

Originally published in *International Journal of Corpus Linguistics*,
<http://doi.org/10.1075/ijcl.22016.ale>)

Abstract

The ways in which politicians have discussed who, what, and where was considered “uncivilized” across the past two centuries gives an insight into how speakers in a position of authority classified and constructed the world around them, and how those in power in Britain see the country and themselves. This article uses the Hansard Corpus 1803–2003 of speeches in the UK Parliament alongside data from the *Historical Thesaurus of English* to analyse diachronic variation in usage of words for persons, places and practices considered uncivil. It proposes new methods and offers quantitative data to describe the period’s shift in political attitudes towards not just the so-called “uncivil” but also the country as a whole.

Keywords: Hansard, Parliamentary discourse, uncivil, barbaric, Historical Thesaurus of English

1. Introduction

Perhaps the worst situation of all is to imagine oneself strong [...] in an actuality of relative decline. Here, we find Britain today: another antiquated kingdom that cannot put jets on its carriers, defeat on its own Gaddafi's ramshackle desert kingdom, nor secure a single province in Afghanistan or a solitary city in Iraq. Yet the UK imagines itself a Great Power, a counter-insurgent nation-builder in far-off lands, a developer of the underdeveloped in Africa and elsewhere. (Barkawi & Brighton, 2013: 1115)

Much has been said about the shifting and complex ways Britain sees itself and its relationship with the outside world, with the situation often simplified to a great colonial empire and world power changing to an uncertain, introspective nation dealing with its postcolonial identity (see, amongst others, Pietsch, 2013; Edgerton, 2018). This long and varied change is disputed by some (Seely & Rogers, 2019; Morris, 2011), but the generally accepted historical narrative is one

of an often unwilling process of change over centuries in British attitudes and perceptions as the UK adjusted to its new place in a changing world. In this article, we look for evidence of this in the Hansard Corpus 1803–2003’s records over two centuries of parliamentary reporting. We aim to account for changes in the discourse of one particular concept which can reflect this change in global context and identity (Skinner, 1969: 50–53), and we do so through a new methodological approach using the *Historical Thesaurus of English* (Kay et al., 2022; hereafter the *Thesaurus*) as a corpus-led historical study in the style of McEnery and Baker (2016).

We do this by analysing the ways in which speakers in the UK Parliament have discussed persons, places, and practices they considered “uncivilized”, from the relevant *Thesaurus* category for that concept. This evaluative construction – being by necessity comparative, subjective, and value-laden (Thompson & Hunston, 2000: 13) – gives us an insight into how speakers in a position of authority classified and constructed the world around them, and how this changed. However, the size of the Hansard Corpus (1.6 billion words) and the polysemy of some of the words involved presents a challenge for standard corpus search techniques.

Therefore, in this article we combine the new lexicological resource of the *Thesaurus* with the Hansard Corpus through the *Glasgow-Lancaster Historical Thesaurus Semantic Tagger*, so as to better investigate diachronic variation in these two hundred years of British Parliamentary speeches. We propose an approach whereby broad diachronic variation in the corpus can be founded on an analysis grouping together distinct lexical expressions, using the *Thesaurus*, into themes in order to develop a coherent dataset of a single evolving concept in English. Our method involves finding ‘sense-family’ groups within the *Thesaurus*’ inventory of means of expressing a concept, retrieving all the uses of that concept in the corpus, and then coding these uses throughout the full range of variation in order to assess relevant shifts in meaning.

In our analysis for this article, the 42 ways to call something “uncivil” in English (as recorded in the *Thesaurus*), can be broadly grouped into five different sense-families, where the people and places who are called uncivil are constructed variously as animals, as crude, ill-formed people, as speakers of a foreign tongue, as savages, and finally as outsiders. In this article, we account for all the evidence from the tagged Hansard Corpus – over 2,000 uses of the *Thesaurus* semantic category 03.01.03.02.01 adj *Uncivilized* – in order to trace across the recorded history of Parliamentary speeches variations and shifts in the cultural, political and social construction of the uncivil. The Hansard Corpus contains records of four of the sense-families above (the recorded use of *the Other* sense-family in the Hansard Corpus is primarily not present, except for one complex instance which we discuss below). Each use of the concept was then semi-automatically coded for a range of values: its membership in each of the four sense-families above; what the referent being modified as uncivil is; the animacy of the referent; if the

referent is foreign or domestic; if it is in the relative present or the past; and the country or people referred to.

Overall, we describe below three primary conceptual shifts found across the two centuries covered by the Hansard Corpus. Firstly, the majority of word uses meaning “uncivil” at the start of the period were being applied to foreign referents, and move to a majority of domestic referents at the end. Secondly, there is a marked shift from discussing uncivil places to discussing uncivil practices. We argue this reflects the increasing levels of verbal oppositionality found in the corpus over the same period. Overall, our new approach to the data shows a coherent pivot from the uncivil primarily being foreigners in the 1800s, to mostly being domestic persons in the early 1900s, and finally to mainly describing barbaric practices and actions in the later 20th century.

2. Our data: the *Historical Thesaurus* and the Hansard Corpus

The SAMUELS project (Semantic Annotation and Mark-Up for Enhancing Lexical Searches, funded by the AHRC and ESRC) produced a semantic tagger based on the work already done at Lancaster University in producing the USAS tagger (Rayson et al., 2004). Building on this work, a team from Glasgow, Lancaster, Huddersfield, and the University of Central Lancashire produced a tagger which used as its dataset the *Historical Thesaurus of English* and could tag diachronic texts at a high level of granularity.

The *Historical Thesaurus of English* (Kay et al., 2022) is a meaning-based rearrangement of the complete contents of the *Oxford English Dictionary*, supplemented by other historical dictionaries.¹ It aims to list every meaning recorded in English along with every word which has been used to realise that meaning, dated with its currency – approximately 800,000 word forms rigorously arranged into over 230,000 hierarchical semantic categories. As well as the only historical thesaurus thus completed, when filtered to data current in the present day, the *Thesaurus* also forms the largest thesaurus of contemporary English currently available, and its data is ideally suited to a wide range of corpus linguistic research.

The *Historical Thesaurus Semantic Tagger* (HTST; see Alexander et al., 2015; Piao et al., 2017, for more detail), produced in the SAMUELS project, takes a corpus of texts and annotates each word with a disambiguated set of meaning codes taken from the *Thesaurus* (in the form of strings of numbers which designate meaning categories). HTST’s disambiguation between the same word forms which relate to multiple senses is built on top of existing NLP tools at Lancaster, including USAS (see above), CLAWS grammatical annotation (Leech et al. 1994), and VARD2 processing of variant spellings (Baron and Rayson 2008, for this project enhanced by variant spellings from the *Oxford English Dictionary*). HTST combines manual sub-lexicons of highly-polysemous words

(such as *take, make, come*) which aid disambiguation, USAS-assisted closed-class mappings (for function words and proper nouns), a context-based disambiguation model based on statistical distance between the HT semantic categories of the word to be disambiguated and its neighbouring words, time-filtering using HT date data (based on the OED), and some other techniques outlined in the papers above. On historical texts of deliberately-high complexity (used to represent the worst-case scenario), HTST agreed with manual annotation between 80 and 90% of the time in disambiguating polysemous words, which is approximately equivalent to manual inter-rater agreement (higher accuracy is unlikely to be gained due to the volume of polysemy in English text and the fine-grained disambiguation needed between word forms with hundreds of possible senses; where HTST and manual annotation did not agree broadly, there was often agreement in general senses). Further work is planned to use databases of metaphorical, extended, and cross-referenced meanings in the HT to enhance HTST, and to evaluate inter-rater reliability of highly polysemous meanings. The tagger is available from UCREL at Lancaster.

This tagger was applied to the Hansard Corpus (originally created by Alexander in 2012) to create a semantically annotated version of the corpus (available online as Alexander & Davies, 2015). *Hansard* is the 1.6bn-word corrected, mediated text of debates in the British Parliament, beginning formally in 1803. We use the 1.6bn-word semantically annotated 1803–2005 version of the corpus in this study (as do other studies involved in content analysis such as Archer, 2017; Hiltunen et al., 2020; Tyrkkö, 2019), although newer versions taking the semantic annotation to 2021 are also now available (see Hansard at Huddersfield Project, 2022; Jeffries & Sanjurjo-González, 2019). Other recent studies have used extracts from *Hansard* to create differently annotated corpora – e.g. Abercrombie and Batista-Navarro (2018a, b) have created a corpus for sentiment analysis using *Hansard* data from 1901 onwards – forming an emerging ecosystem of *Hansard* corpora. From our own corpus, we retrieved KWIC entries for all the occurrences of a particular *Historical Thesaurus* semantic category – that of the *Uncivil* (see below) – and manually checked in the corpus for other occurrences of the same words to check if any had been mistakenly tagged, at which point they were added into the database.

A full discussion of the editorial practices of *Hansard* editors would take up more space than is reasonable in this section (see Alexander, forthcoming), but the corpus is a large, formal, and highly accurate account of debates *in terms of topics discussed*. No single mode or genre of language is ever perfectly representative, and Parliamentary discourse is a genre which corpus linguists perhaps unfairly treat with more suspicion than most. Parliamentary reports are highly edited, and writing *Hansard* involves keeping the substance of what is discussed while erasing infelicities which would detract from the sense when recorded in writing. This means that it is wholly unsuitable as a transcript if one is interested in preposition use or hesitation phenomena, but one of the best

records available when it comes to the primary sense-carrying parts of speech such as nouns, verbs, and adjectives. Much of it consists of written-to-be-spoken material, and it is often influenced by ideals of formal rhetoric as well as the local genre conventions of Parliamentary interaction. While the debates recorded here were shaped by the political, social and cultural climates in which the speakers took part (Skinner, 1969: 53), the precise patterns of speech were also shaped by the culture within Westminster. It is important to note, therefore, that while “the Commons itself is usually treated simply as the venue for political activity”, in actual fact “controversies over Parliamentary style were indicative of more general disputes over the culture and ideology of British politics” (Toye, 2014: 271). We therefore do not focus on phraseological phenomena (or even collocational phenomena), but singularly on word choice: these decisions, conscious or not, which are used to detail politically and ideologically-driven concepts inform us of the ways in which politicians construed such contested categories.

One particular concern of using *Hansard* data is the varying transcription methods and policies in use across the full span of the period. As analysed in Alexander (forthcoming), from 1803-1876 speeches were reported primarily in the third person with major speeches (from key ministers or on important topics) reported in the first person, in full. This expanded in the periods 1877-1888, 1892-1897, and 1899-1908 to include more reporting on “unpopular” topics due to subsidy. In 1889-1892 and 1898-1899, *Hansard* consists of first-person shorthand reporting, with routine speeches occasionally compressed somewhat. In the period 1909 onwards, all speeches were reported in full from shorthand (and later from recordings). As a result, it is not entirely accurate to say that 1909 marked a complete change of the text-type of *Hansard* – the text type is actually highly variable throughout the first century or so of its existence – although it is the case that there is a shift in both the stylistics of the report and in the volume of column space given to speeches thought at the time to be unimportant. The shift in style is a concern for many studies – although not as much of one in studies of open class words such as this article undertakes, which are less affected by third-party reports as other lexical entries are – and the rapid shifts in the 1888-1909 period between different transcription methods provide a useful check as to how much the third-person reportage style alters the frequency of use of these words, where our analysis of our set of words shows little variation. The “unimportant” speeches underreported were generally in public or financial committees and on private bills (which do not affect the general public but which, for example, authorise individual marriages or land transactions) – we know this because Parliament for a period ended up subsidising the recording of particular categories which were underreported, and for the purposes of this article such discussion types are less likely to involve foreign or domestic relations of the state. While we are highly conscious of the underlying transcription issues of the corpus, they do not to us appear to affect the study we undertake here. For

example, the clear decade-on-decade trend in Figure 6 continues despite the large internal variation in transcription practice across the decades 1870–1920. It is also for this reason that we do not use standardised frequencies in some figures below: *Hansard* increases in size over the years and so a per-million standardisation may seem attractive, but its increases in size will be in non-comparable material (it is not presently possible to distinguish throughout the Hansard Corpus only the material on public bills and government foreign relations, the nouns and adjectives from which would be a useful initial standardisation measure, with a modifier applied to the multiple phases of shorthand transcription). In the visualisations below, raw counts – approached with a knowledge that they are counts – are less misleading than standardised frequencies due to the nature of the corpus. Similarly, other corpus linguistic uses and methods applied to the data, such as studies of speech types, interactional phenomena, or stylistics, should be undertaken with caution as advised in detail in Alexander (forthcoming) and initially outlined by Slembrouck (1992) and Mollin (2007). The Hansard Corpus’s long age, appealing content, and surface homogeneity make such warnings notable, although in context they are not more substantial than that for any large corpus.

The combination of the precise semantic data of the *Thesaurus* and the annotation provided by HTST allows searching of the Hansard Corpus by semantic field. Below, we explore a specific *Thesaurus* category of words – 42 adjectives – in the Hansard Corpus, which can be searched for in the online corpus by semantic category to return only those words which the tagger has identified as being within that semantic category.

3. Sense-Families and the Uncivil

Writing in 1900, the Secretary of the Aborigines Protection Society, H. R. Fox Bourne, noted that “the number of uncivilized people brought into contact with nations calling themselves civilized has been enormously increased” (Bourne, 1900: 4). The history of the speakers of English involves ongoing contact with those who were deemed to be “uncivilized”; the varied lexicalization of this concept displays the complex attitudes of those who exercised meaning-making and ideological power throughout the last two hundred years. Many attitudes revealed in this way were unconscionable then and are unconscionable now, and yet the language used is deeply intertwined with the ideological positioning of a ruling political class throughout the expansion and contraction of the British Empire. Such linguistic realizations of dominant and naturalized ideologies offer a window onto the socio-political construction of those persons artificially placed into an inferior position (Kumar, 2012: 76–101).

There are shifts in sense and usage over time of these linguistic realizations of the “uncivilized” (we will not continue to use these scare quotes,

but the term should be understood to always include them). In previous work, we have put the 42 adjectives recorded in the *Thesaurus* category of *Uncivilized* (03.01.03.02.01 in the *Thesaurus* and AY:03:b:01 in the simplified HTST corpus annotation scheme paralleling *Thesaurus* categories) into five groups we call ‘sense-families’ (Alexander & Struan, 2013). These families are groups of words which share a common construal of the uncivil, and so instantiate a similar worldview, and we outline them below. The approach of grouping words in this way allows us to more easily investigate and analyse usage within a particular *Thesaurus* category over time, and in so doing trace the development of particular concepts. Throughout this article, we use the term *Uncivilized* with a capital letter to indicate the concept, with all its various lexical realizations, and to distinguish it from the lexical entry *uncivilized*. The same principle applies to the sense-families discussed below.

When looking at the category of *Uncivilized*, we see a path of lexicalization that reflects the times within which the words were coined. The adjectives recorded here cover the period of English from Old English through to the modern day, and within them we see terms of varying levels of ideological construal, such as *wild*, *bestial*, *barbarous*, *irreclaimed*, *incivil*, *medieval*, *savage* and *pre-civilized*. What we now consider to be the dominant term – *uncivilized* – was a relative latecomer to the category, with its first recorded use in 1607. The contents of the *Thesaurus* category for *Uncivilized* (category 03.01.03.02.01, adj.) can be visualized in Figure 1.

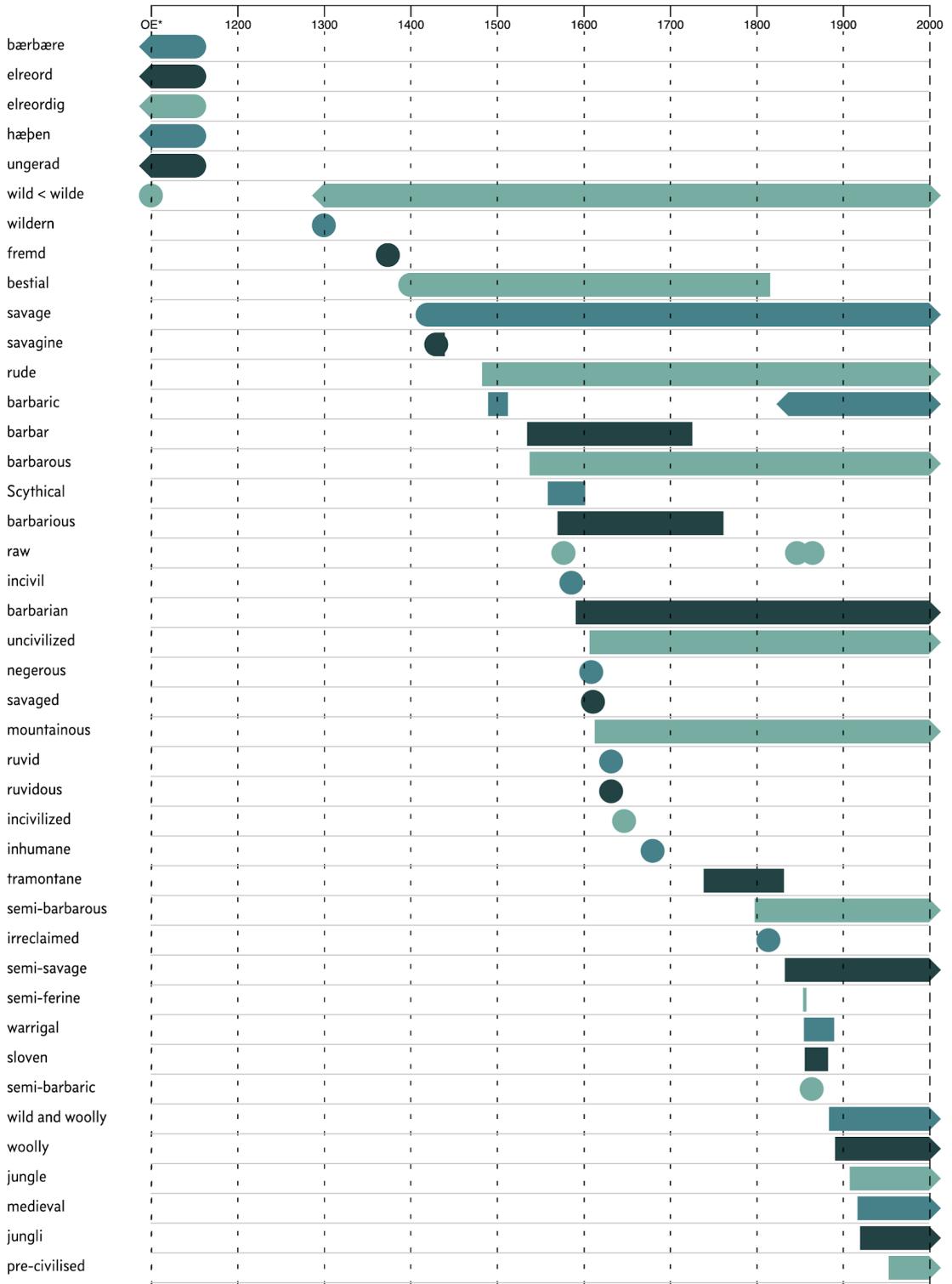


Figure 1. The *Uncivil* category in the *Thesaurus*, visualised to show length of attestation of each word

The timeline in this figure uses bars to represent stretches of time and circles for single dates. A flat ending to a bar is a precise date which begins or ends a word's currency, while a curved end indicates an approximate date. A pointed end to a bar indicates the word's use may extend beyond that period, such as the indicators of continued modern currency to the right of the visualization. While there are five words only known to exist in the Old English period, twenty words have a significant period of currency, including sixteen still in modern use.

The five sense-families within these words which we previously described have developed in their usage as the speakers of English came into contact with an increasingly international, diverse world, and as the concept of civility itself shifted in meaning to suit the prevalent worldviews of the time. The two eldest of the sense-families we have grouped under the headings *Wild* and *Crude*. The third sense-family we call *Barbar*, including the variations around the concept of "barbarous". The fourth sense-family is *Civility*, and it begins and spreads as the modern concept of the state developed in Western Europe. The final sense-family we call *the Other*; in this sense-family, we see Othering of groups based on racial, geographical and temporal location. All sense-families of *Uncivilized* are available to the modern speaker; for the period covered by *Hansard* (the final two columns of the visualization above), speakers of English would have had a wide variety of terms to choose from when speaking of peoples, places, practices and things they found to be uncivilized.

To be *Wild* is to be animalistic and without humanity. To be in this category, then, was to be out of society by being *bestial* or *savage*. Speaking in the 1860s, the Liberal MP for County Louth, Chichester Fortesque, used the term *savage* to describe the peoples with whom British colonists came into contact; see Example (1).

- (1) ...he found islands which comprised the most extraordinary collection of races that had ever been brought together under any Government; and a number of dependencies with savage tribes on their frontiers or within their borders.
(s3v0158p0_03203, 31/05/1860)

These so-called savage tribes were outside the realms of civilized society, and crossed borders without hindrance; for example, in Southern Rhodesia, "a number of savage alien terrorists to invade [...] British territory" (s5LV0402p0_02256, 13/11/1979). For the words in the *Wild* category, the terms provide an animal cruelty and a pejorative discussion of those without even the base-level hallmarks of civility.

The second of the sense-families developed the concept of being animalistic to being unfinished, unpolished. This sense-family – which we label *Crude* – includes words such as *rude*, *raw* and *rough*. The people of Central Africa could, therefore, be described as being "rude in their government and policy" (s3v0245p0_00146, 31/03/1879). The use of terms in this category is to

emphasise what a people do not have; in other words, their incivility is marked by their incompleteness and their roughness. This sense-family has few recorded uses in the Hansard Corpus and remains the smallest of the sense-families.

By the late Medieval period, speakers of English had adopted a new dominant term to categorise those deemed to be uncivilized; this we have categorized as our third sense-family of *Barbar*. This term, originally from the Greek *βάρβαρος*, at first denotes speakers that sounded different from Hellenic Greeks. Its first recorded use is in 1535, and variations around the term *barbar* have existed in English since that time. The original connotations of difference in sound remained for a short period in English but were soon overtaken by the overarching concept of being without civility (we discuss this further in Alexander & Struan, 2013). To be barbarous was to be without civility, polish or the finer elements of society: “It had been well observed, that the possession of iron was one of the great grounds of distinction between civilized and barbarous society” (S1V0007P0_18060509, 09/05/1806). More than that, however, to be a barbarian was to be fully without care or due respect for mankind: “Her Majesty's Government deplore this barbaric behaviour. Such actions are certainly in defiance of the spirit of the Universal Declaration of Human Rights” (S5CV0566P0_00470, 04/03/1957).

Barbarism was, however, something which could be escaped; by transferring society, action or practice into civilized manners, formerly barbaric peoples, practices and places could become civilized. Speakers of English classified their ancestors as barbaric: “We boast that we have abolished our ancient and barbarous mode of making extensive revolutions and changes in our political system” (S3V0153P0_00999, 24/03/1859) and “in fact, the Bill itself seemed to him to be a return to the ancient barbaric days. It was an endeavour to harmonize subsisting institutions with those existing in the Anglo-Saxon times” (S3V0206P0_00946, 12/05/1871). This view, of a society pulling themselves out of barbarism through the creation of a constitution (written or unwritten) is a common theme in Anglophone politics (Colley, 2014: 249-252).

The ability to remove oneself from barbarism resulted in the concept of being semi-barbarous (in recorded use since 1798). A people could remove themselves from their barbaric state by adopting civilized norms and practices: “With regard to Russia, he professed to know nothing. They might be a barbarous, a semi-barbarous, or a civilized people, as they had been variously represented” (S1V0024P0_00760, 30/11/1812). The ability to continue in this state of semi-barbarity remained to the modern times: “We can now see that in dealing with Russia we are dealing with a semi-barbarous state and a society that only knew a measure of democracy for a few years before the First World War” (S5LV0563P0_02159, 18/04/1995).

The development of the early-modern and modern state created in English what we term *Civility*, our fourth sense-family. As mentioned above, the term

uncivilized entered recorded use in 1607, and is used to confer status within or without the accepted norms of a “civil society”. “Whenever we came into contact with men with black skins we called them uncivilized barbarians, when the truth of the matter was that they had the misfortune to own lands which we coveted” (S3V0176P0_02968, 14/07/1864). These so-called *uncivilised* peoples could be brought towards civilization as part of the British Empire’s “civilizing mission”: “That is, I think, a great danger in the case of an uncivilized people like them; but it is impossible not to hope that their knowledge of the numerous blessings and benefits which they have derived from being under British government” (S3V0243P0_02649, 13/02/1879).

The concept of *Incivility* differs from the categories above: the classifications above all rely on a process of Othering based on various characteristics. For *Incivility*, however, the concept of Othering was based predominantly on the relationship with the state: “Malta had been governed as a Crown colony, as though it were a conquered or uncivilized country, and not as being a people who voluntarily placed themselves under British protection, believing that their ancient rights and privileges would be confirmed to them as British subjects” (S3V0318P0_01714, 29/07/1887). This term posed an ongoing difficulty for European thinkers, and especially for liberal critiques of empire-builders. The role of the state in the process of colonization, and in the processes of a supposed civilizing mission, proved an ongoing debate for the British (Fitzmaurice, 2012: 122-140; Bell, 2010: 34-64).

The final of the sense-families, *the Other*, categorizes others as racially, ethnically and geographically distinct and different. As the most recent sense-family, this is not found in the Hansard Corpus except in the word *medieval*, in the sense of a practice or people so severe, cruel, uncivilized, or “illiberal” (as the *OED* definition puts it) that they reflect the stereotypes of that age. This new sense of the word (emerging in the twentieth century) poses problems for our analysis, as it has complex metaphorical underpinnings: a mention of a “medieval administration” could mean the administration is seen as uncivilized, or it could mean cruel, violent, or backwards in some other manner. We have reserved the word *medieval* and its uses in this context for future research, in order to investigate its figurative extension more thoroughly and explore the connections between the concepts of *Uncivil* and cruelty, which *medieval* (and to a much lesser extent barbarous) bridges. This, however, should not affect the discussion below, as a series of manual analyses on random 100-occurrence samples of the use of the word *medieval* has shown that only 2% on average of the uses of that word would be candidates for meaning *Uncivil*. Therefore, *the Other* sense-family does not appear below.

The sense-families discussed above give us the ability to classify the various conceptions of incivility we find in our corpus. When used as an analytical tool, sense-families therefore permit us to trace changes over time of the five distinct phases of the concept of incivility with nuance.

4. The Development of the Uncivilized

The situation of Great Britain brings her beyond any other power into communication with the uncivilized nations of the earth. We are in contact with them in so many parts of the globe, that it has become of deep importance to ascertain the results of our relations with them, and to fix the rules of our conduct towards them. We are apt to class them under the sweeping term of savages, and perhaps, in so doing, to consider ourselves exempted from the obligations due to them as our fellow-men. (Select Committee on Aboriginal Tribes, 1837: 1)

For centuries our country has looked outwards. Today the links between countries and communities are closer and more immediate than ever before. As the barriers of the Cold War have crumbled, a technological revolution has been erasing obstacles of time and distance. [...] But globalisation – including rapid economic growth and greater freedom of movement – also brings risks. Some parts of the world – Africa and the Middle East in particular – are in danger of being left behind. [...] The world today is different. The opportunities of global economic change are not the preserve of an elite few. (Foreign and Commonwealth Office, 2006: 4)

The UK's relationship with the outside world has undergone substantial change over the past two centuries. Writing in the 1837 report from a Select Committee of the House of Commons (tasked with investigating how the British government interacted with, and in their view protected, the various peoples in contact with the British Empire), we see Britain at the height of her power attempting to systematize approaches to the various colonized peoples. By the start of the twenty-first century, however, the British view of the world – and the world's inhabitants – had changed dramatically, and the Foreign and Commonwealth Office (FCO; 2006) report highlights this shift of perception of those outside the United Kingdom's borders. The two centuries covered by the Hansard Corpus document the rise and demise of the British Empire, over a period of significant change for speakers of the English language. These changes are reflected in the ways in which British politicians – speaking in both Commons and Lords – framed peoples, places, practices and actions as uncivilized.

The unique combination of the *Thesaurus* with the Hansard Corpus allows us to investigate the 1,968 instances of the use of any word within the *Thesaurus* semantic category of adjectives for *Uncivilized*. As discussed above, we also further categorise these words into sense-families (see Alexander & Struan, 2013). In our analysis of all the uses of the terms, we identified and further

categorized the referents of each term – are they nations, people, places, times, and so on – as well as whether the use of the term was in reference to the present or the past, and foreign or domestic.

4.1 Types of referents

The first way in which we can see the shift in ideology is through what the *uncivil* adjectives are being used to describe. We further categorised these referents into the following eight categories:

- (i) nations
- (ii) object/thing
- (iii) people (this can be referring to a specific people, but may not be what is considered a full nation or nation-state)
- (iv) place
- (v) place, time (this is a specific place at a specific time, so a defined location at a given point in time)
- (vi) practice/action
- (vii) state (in the sense of “a state of barbarism”, not to be confused with a nation-state)
- (viii) time

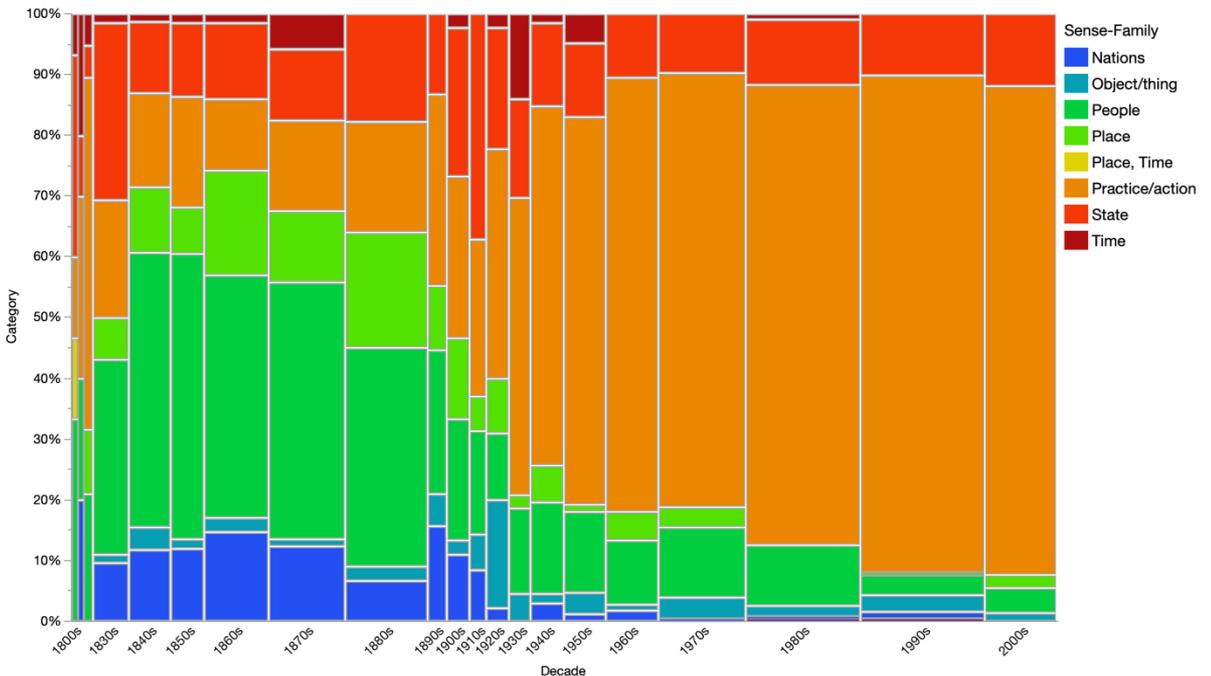


Figure 2. Mosaic plot of types of noun referents of the *Uncivil* by decade

In Figure 2, we can see the change over time in the types of nouns being modified by *Uncivilized* adjectives. The plot shows, for each decade of the corpus, what percentage of the nouns modified by *Uncivilized* are in each category – the thickness of each decade reflects how many occurrences there are in each decade (with an approximate average of 100 hits per decade), meaning the size of each category is meaningful in both the horizontal and vertical dimensions.

The changes in usage throughout the two centuries are stark: British politicians throughout the 1800s refer overwhelmingly to people, places, and nations as being *Uncivilized*, while their late-twentieth century counterparts make almost no reference to these, with practices becoming overwhelmingly almost all that is considered as *Uncivil* as we reach the present. Politicians could therefore speak in the 1830s of “barbarian India” (s3v0032p0_01148, 13/04/1836) and that colonizers could “promote civilization in the barbarous parts of Africa” (s3v0088p0_00299, 31/07/1846). Some of these specific referrals remain in the second half of the twentieth century (“I was in that barbaric country, the Yemen, some few years ago” (s5lv0306p0_02480, 17/12/1969)), but the overall number of references drops off substantially and most place or nation references instead refer to specific characteristics inherent in some regimes – for example “how can he agree to the holding of [the 1980] Olympic Games in one of the most barbaric and totalitarian countries in the world [the USSR]?” (s5cv0975p0_03470, 12/12/1979) and “the death and torture camps; the barbaric prisons for political opponents” (referring to Iraq; s6cv0401p0_02799, 18/03/2003).

The first primary shift in the use of words for the *Uncivil* in these two centuries of Parliamentary discourse therefore is a clear movement from principally describing people and places to describing practices. These practices in later years can frequently still be foreign, but they are somewhat less face-threatening in their use, and they show emerging precision and discernment in the application of the label *Uncivil*: no longer a widely-cast blanket term for all people in a race or country, it is instead particular phenomena (prisons, deportation, terrorism, foxhunting) which are judged and found wanting.

4.2 Sense-Families

We can see other patterns of change in usage of particular sense-families over time. Figure 3 is a heatmap which highlights the prevalence of use of terms based on the four represented sense-families. We see that the eldest categories – to be *Wild* and *Crude* – are minimally used in Parliamentary speech, and when they are used it is primarily for rhetorical effect, with purposeful effect and emphasis; for example, the depiction of a settler company in the Gold Coast as being the “singular office of dispensing rude laws among uncivilized tribes”

(s3v0177p0_02082, 21/02/1865) or to describe “those who were attacked by the Nazis in such a bestial way” (s5LV0361p0_04757, 26/06/1975). By the time of the *Hansard Corpus*, these two early ways of referring to the *Uncivil* have become either out of fashion or insufficiently nuanced that they are primarily reserved for such flourishes rather than political discussions.

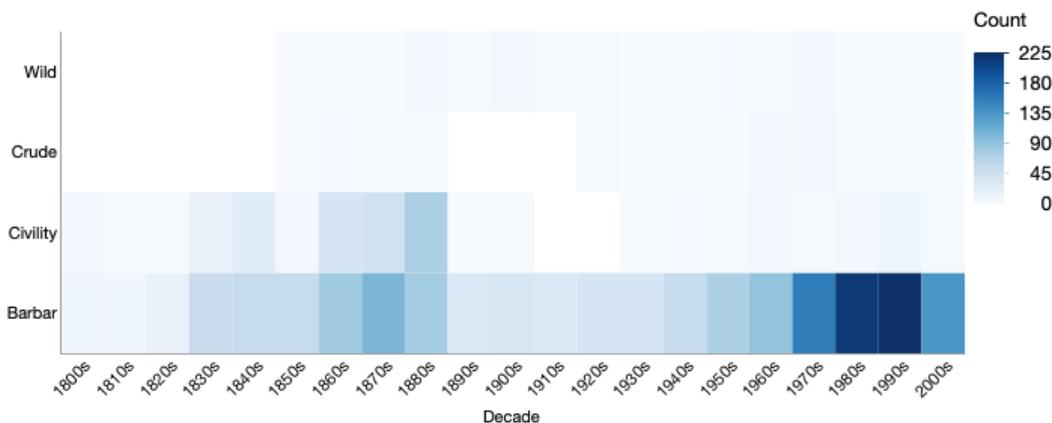


Figure 3. Heatmap of sense-family use by decade (note these heatmaps use unmodified count figures of word use; see the discussion in Section 2 above on this)

Civility, a term in fairly frequent use throughout the nineteenth century, drops in use after the end of the 1880s, with no recorded uses between 1909 and 1933 (s5CV0005p0_04238, 27/05/1909 and s5CV0280p0_06547, 24/07/1933), and only 38 uses from 1933 onwards. In its place, we see the already-popular *Barbar* sense-family rise in use, with a very pronounced rise from the 1970s.

Figure 4 further highlights the dominance of the *Barbar* sense-family within the Hansard Corpus. This heatmap shows the use of each word per decade, and shows how much of the data is made up of recent uses of the word *barbaric*; uses of *barbaric* since the 1930s alone account for 822 of the 1,968 occurrences in our data.

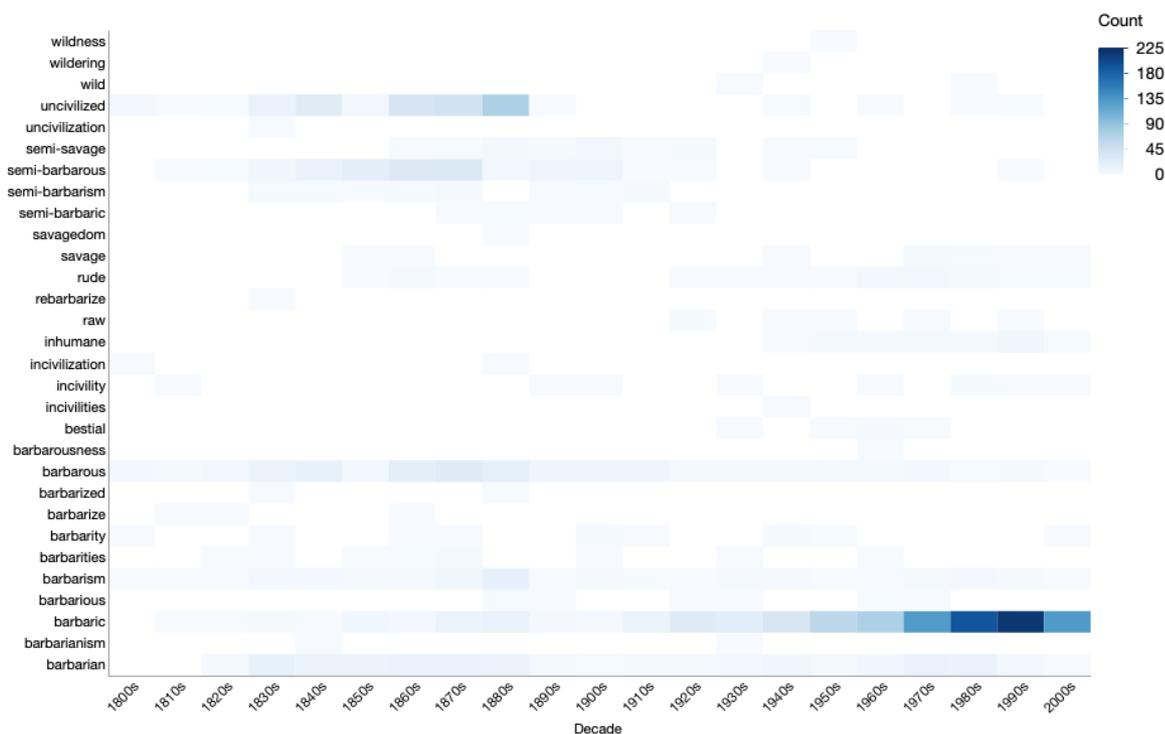


Figure 4. Heatmap of lexical items by decade

What makes *barbaric* so attractive? As well as its pleasing phonetic shape, the word is usefully polysemous for a speaker, particularly one focused on rhetorical effect: it can mean “cruel” and “harsh” as well as *Uncivilized*, and so carries with it additional pejorative meanings than simply being incivil or savage (which can be a state of innocence or nobility rather than baldly negative). Its clear associations in modern English are seen by its top ten BNC (1994) collocates, which include *practice*, *cruel*, *punishment*, *bestial*, and *savagery* (there is an occasional use of the fixed phrase *barbaric splendour*, which on the surface appears oddly complimentary but is a term principally used by colonizers in a rather condescending fashion). While *barbaric* may be the most commonly used word for *Uncivil*, it is far from the most neutral available in modern English.

The word *barbaric* also has a significant association with practices and actions, as shown in Figure 5, which only displays the use of *barbaric* in the categories discussed in the above subsection. Its most common collocate by far is *practice* (collocating 121 times out of 1050 occurrences), closely followed by *cruel* (50), *acts* (44), and *practices* (33). These collocates and Figure 2 above show that the increase in *barbaric* can be explained – at least in part – by its close association with describing practices and acts. In *Hansard*, as elsewhere, *practices* has a strong negative semantic prosody, meaning *barbaric*’s pejorative rhetorical effect is useful for a speaker in combination with *practices*. We return to

this below, but the increase in emphasis on describing practices as *Uncivil* is one which promotes the selection of *barbaric* more often out of a speaker’s options.

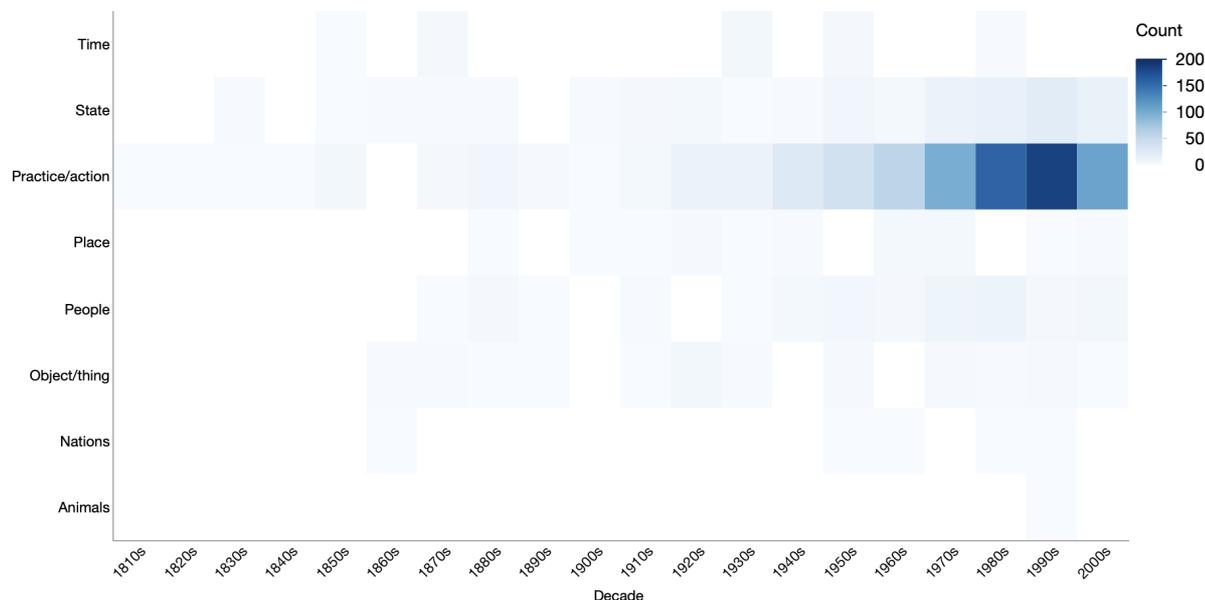


Figure 5. Heatmap of types of noun referents for *Uncivil* terms

4.3 Foreign and Domestic

Each occurrence of the *Uncivil* adjectives was coded for whether it referred to a foreign or domestic referent, as shown in Figure 6. The complex situation of the island of Ireland as it relates to the British Parliament means that Ireland and Northern Ireland were coded separately to the domestic and foreign referents, which otherwise straightforwardly referred to referents within and outwith the UK, respectively. By separating Ireland and Northern Ireland, we aimed to avoid the issues of identifying whether the speaker was conceptualising Ireland as a foreign or domestic entity, especially given Ireland’s contested history of representation in the UK Parliament. The *Uncivil* data in *Hansard* has the potential to contribute further to our understanding of the conceptualisation of Ireland in the minds of politicians, but in this article we restrict ourselves to the foreign and domestic categories.

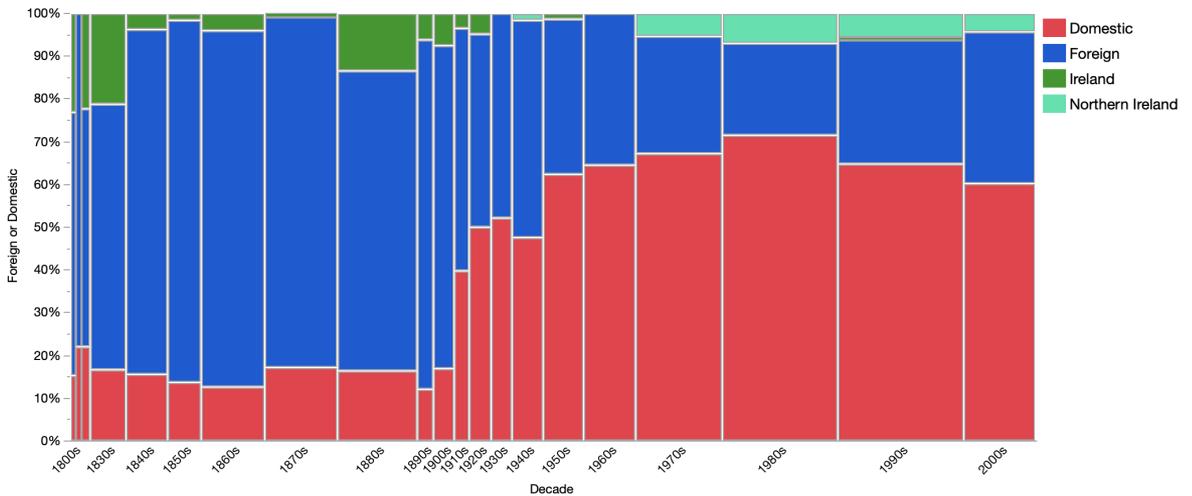


Figure 6. Mosaic plot of foreign/domestic referents by decade

Of our data, the top twenty locations identified as being *Uncivil* are shown in Table 1. The most frequent *Uncivil* location(s) is/are Ireland/Northern Ireland, followed by India, Russia, Scotland, Africa, and China. Setting (Northern) Ireland aside, there is no strong pattern for the most common country referents – India was colonized by the British, Russia was not, and Scotland is a domestic referent. Of the ten main *Uncivil* locations after Ireland, from India to Europe in the table, five were colonized by Britain (India, Africa, New Zealand, South Africa, and Iraq, although Iraq’s mentions are primarily in the period of the Gulf and Iraq Wars, from 1990 onwards).

Table 1. Top twenty countries referred to as *Uncivil*

Ireland	55
Northern Ireland	45
India	34
Russia	32
Scotland	31
Africa	30
China	25
New Zealand	18
South Africa	16

Turkey	16
Iraq	14
Europe	13
Abyssinia	11
Japan	11
England	10
Greece	10
Spain	10
Afghanistan	9
US	9
Egypt	8

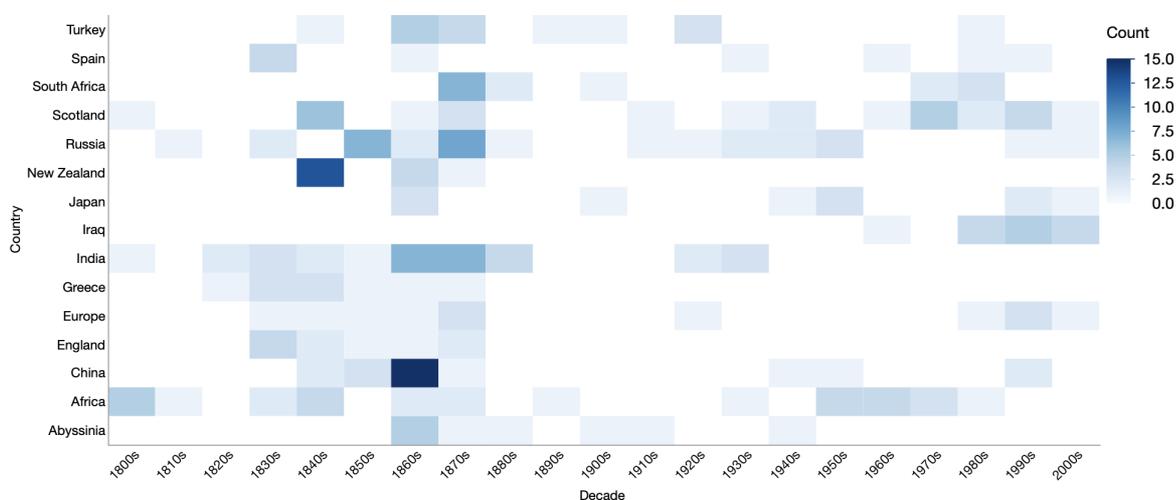


Figure 7. A heatmap of countries referred to ten or more times as being *Uncivil*, by decade

Figure 7 takes all the locations with ten or more *Uncivil* references as a heatmap of country references across time. The heatmap in Figure 7 highlights the ways in which British politicians have engaged with and categorized the external world. South Africa’s foundation as a self-governing colony, Cape Colony, in 1872 resulted in significant debate in the British Parliament: “You have to deal in South Africa with a vast Native population within the Borders of your Colonies –

a population in every state of civilization, varying between the civilization of the Whites and actual barbarism” (S3V0244PO_02954, 27/03/1879). *Africa* as an *Uncivil* term is used across the nineteenth century, and during what historians call the Scramble for Africa, primarily to refer to the whole continent during a significant expansion of colonial imperialism, and then during the period of direct colonial rule mentions of Africa as a whole decrease (with mentions of individual colonies and nations increasing, as politicians become more selective in their *Uncivil* referents). The decolonization period then sees rises in the mention of Africa as a whole again, once more grouping the continent together as a whole *Uncivil* entity.

Similarly, the peak of discussion around New Zealand in the 1840s was as the British negotiated the Treaty of Waitangi (1840), declared New Zealand a British colony, and began to govern the islands: “He did not ask them to express an opinion whether the Treaty of Waitangi, or any other treaty between civilized and barbarous nations, was wise or foolish” (S3V0174PO_02877, 26/04/1864). We see, too, in discussion of India a spike around the creation of the British Raj (1858–1947). Parliament’s reaction to the process of Crown colonization of India resulted in ongoing debate and discussion around the ways in which the peoples of India could be cast as different to, or lesser than, their British counterparts, as illustrated in Example (2). The process of colonization, then, resulted in increased discussion of the ways in which the Parliamentarians at the heart of successive British governments codified and categorized the colonized as *Uncivil*.

- (2) India had for ages been celebrated for its civilization. India had been celebrated for barbaric splendour; but he did not believe that in this country, at all events, gorgeous palaces and splendid trappings were regarded as marks of civilization. All the principal monuments of India were raised to gratify the taste or vanity of despotic monarchs. (S3V0156PO_02288, 09/02/1860)

The heatmap also allows us to see discussions around particular crisis points in British foreign policy: the Opium Wars with China (1839–1842 and 1856–1860); increasing Russophobia through the 1870s (and the marriage between the two countries’ royal dynasties); and the Iraq wars at the end of the twentieth century. In addition, when grouping the referent categories in 4.1 above alongside if they have foreign or domestic referents, we come to Figure 8, which shows in the first eight rows the categories which apply to foreign referents, and then in the next eight rows those categories which apply to domestic referents. Mirroring the findings above, it is clear that a general preference in the nineteenth century for foreign people as *Uncivil* referents has shifted to one for domestic practices. There is an increase in referents of foreign practices over the period, but it is substantially outnumbered by the dramatic increase in domestic practice referents. We return to this below.

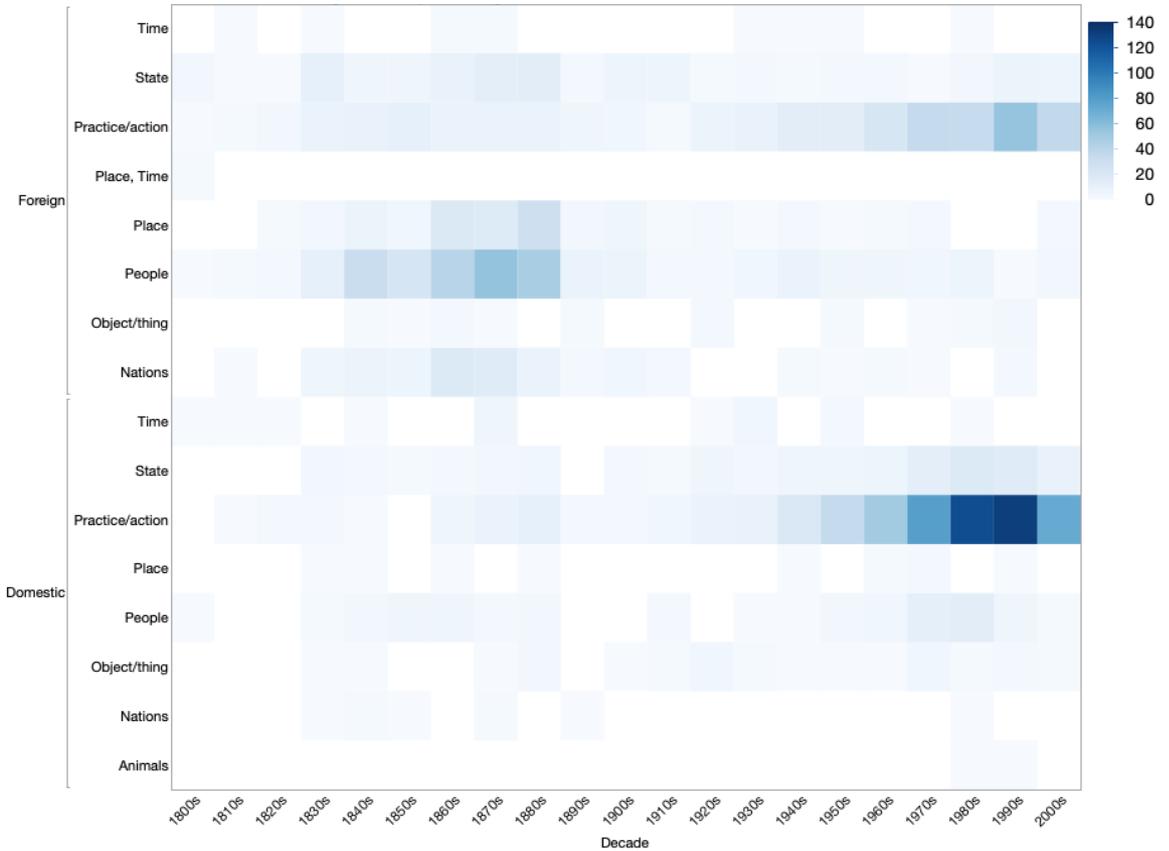


Figure 8. A heatmap of categories of *Uncivil* referents, grouped by foreign/domestic, by decade

The change in focus between foreign and domestic referents means that we could have, for example, discussion of “Nicholas of Russia, with all his barbaric pomp and power, dared not attempt a tenth part of the tyranny of Louis Philippe [of France]” (s3v0031p0_01385, 05/02/1836) and “the half naked and uncivilized chiefs of Africa” (s3v0080p0_00346, 05/05/1845) in the nineteenth century. By the latter parts of the twentieth century, however, the focus of the discussion had shifted to domestic practices: “Nothing is righteous about such barbaric killing [of poultry]” (s5cv0829p0_00665, 17/01/1972) and to ask “how it was possible for hunting with dogs to continue into the 21st century: It is a barbaric practice, and it must come to an end” (s6cv0424p1_20040915, 15/09/2004).

4.4 Past and Present

Finally, there is a distinction between if the referent of an *Uncivil* adjective is in the present or in the past, relative to the speaker. In this respect, the data as

shown in Figure 9 does not show as marked a shift as above, but there is a distinct trend – at the beginning of the nineteenth century, on average approximately half of occurrences were about the present being *Uncivil*, with the other half being about the past – with some variation in the late nineteenth century and early twentieth, with the Second World War understandably marking the high point of present *Uncivility* – the second half of the twentieth century settles into a fairly stable two-thirds present, one-third past pattern. The shift is not very substantial, but does show that the present is increasingly what is considered *Uncivil*.

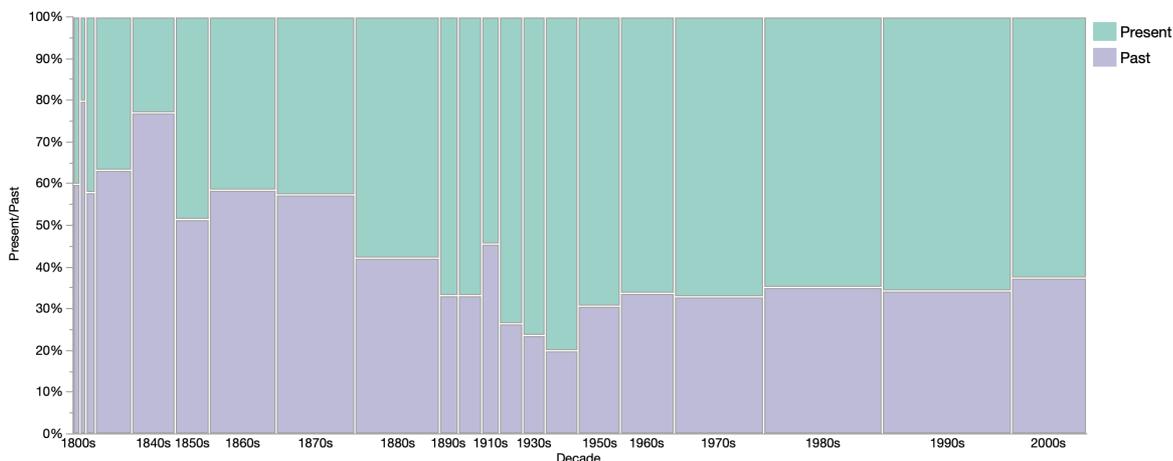


Figure 9. Mosaic plot of present/past *Uncivil* referents by decade

5. Conclusions

The *Uncivil* data above shows the following trends over the two centuries of the Hansard Corpus. There is a clear movement from *Uncivil* describing mainly people and places to describing mainly practices. This shift to practice-based referents then heavily promotes the use of the word *barbaric*, with its more strongly pejorative associations than many other words in the *Uncivil* category. In addition, during this period there is a pronounced change from *Uncivil* being used for foreign referents to domestic referents. Finally, there is a much smaller shift to have proportionally more referents referring to the relative present than the past. Taken in tandem, we see from these a unified semantic drift from *Uncivil* being associated with places and persons in foreign lands at the start of the corpus to in modern times being associated with domestic practices, primarily in the present day and with a relatively high level of pejorative connotations. These changes are most pronounced from the 1950s onwards, at the same time as the British decolonization of Africa, Asia, and the Americas, but the trends are present across the long timespan of the corpus.

We began this article by describing the historical assertions of how Britain, in a changing world, rose and declined in importance and power across the last two centuries; this is perhaps the defining feature of the last century or more of British political history. Not all scholars agree, and some modern politicians – depending on their ideologies – would argue the point. Often the argument is made impressionistically, based on personal experience, perceptions, or surveys. Away from this style of argument, however, the *Uncivil* data we have assembled demonstrates the move in Parliamentary discourse from an Imperial condescension used to describe nations and peoples towards a self-critical approach to modern domestic practices. The changing context of the world outside the Houses of Commons and Lords affects the choices made in speeches within it; whether a politician would be comfortable with the facts or not, the decline in Britain’s relative power is clearly shown in a decline in its Parliament openly and immodestly evaluating the outside world.

The increasing use of the more highly pejorative term *barbaric* in the domestic arena could be seen as a rise in harsher evaluation in recent years, paralleling the evergreen claims of a degradation in political rhetoric in recent years. The high use of *barbaric* at the same time as an increase in domestic focus of *Uncivil* terms, however, does not straightforwardly reflect an increase in negative self-reflective evaluation, but rather can be seen as a rise in oppositionality. The *Thesaurus* category *Harsh/severe* (01.15.21.05.03 adj., AO:22:d:03) i increases in almost every use of highly pejorative words in the later years of the Hansard Corpus, including *draconian*, *atrocious*, *harsher*, *tough*, *harsh*, and *dangerous*. It is outwith our scope to describe this rising-stakes rhetoric, but the increase in *barbaric* in describing domestic practices is not one we would interpret as necessarily a reflection of a change in evaluative discourse at the same time as the changes we describe above, but rather one of a generally increasing preference for the selection of more strongly pejorative lexical items.

The semantic drift of the *Uncivil* in this corpus shows a shift of focus in Parliamentary attention not easily seen through standard histories of the legislature. As the place of Britain in the world has changed, the place of these adjectives has also moved in use: an outwardly judging nation becomes an inwardly judging one in a way which is often difficult to pin down. The category of the *Uncivil* is an effective lens with which to analyse this specific characteristic of modern political discourse, and it is one which could not have been brought to bear without the confluence of a set of new methodological approaches and frameworks. Without the fifty years of painstaking assembly of historical synonyms represented within the detailed hierarchy of the *Thesaurus*, we could not effectively gather the available varied realisations of this complex concept. Without semantic tagging and powerful corpus search engines, we could not reasonably explore the 1.6 billion words of *Hansard*. And without new methods (such as sense-families) to group, categorise, and analyse the results, we could not meaningfully interpret the shifts in discourse over such a long period of time.

The combination of all three is particularly powerful, even when analysing the conspicuous loss of power.

Acknowledgments

This work was supported firstly by the Arts and Humanities Research Council in conjunction with the Economic and Social Research Council (grant AH/L010062/1) in corpus creation and in the creation of the HTST tagger, and latterly by the Leverhulme Trust (grant PLP-2019-341) in the analysis and write up of this article. We acknowledge the assistance of Elina Koristashevskaya in the initial data categorization. The title quote is taken from the corpus, and is from a speech by the Earl of Carnarvon on 12 May 1874 in the House of Lords.

Notes

1. At present, as well as dictionaries of Old English, this means the first edition of the *Oxford English Dictionary*, together with the two later supplements (all of this being approximately equivalent to the second edition of the OED in 1989), the 1990s' three-volume OED *Additions* series, and new words from the in-progress OED3. See <https://ht.ac.uk> for more details.

References

- Abercrombie, G., & Batista-Navarro, R. (2018a). A sentiment-labelled corpus of Hansard parliamentary debate speeches. In *Proceedings of ParlaCLARIN*. Common Language Resources and Technology Infrastructure (pp.43–47). CLARIN. http://lrec-conf.org/workshops/lrec2018/W2/pdf/book_of_proceedings.pdf.
- Abercrombie, G., & Batista-Navarro, R. (2018b). *HanDeSeT: Hansard Debates with Sentiment Tags* [Data set]. Mendeley. <https://doi.org/10.17632/XSVP45CBT4.2>
- Alexander, M. (forthcoming). Speech in the British *Hansard*. In M. Korhonen, H. Kotze, & J. Tyrkkö (Eds.), *Exploring Language and Society with Big Data: Parliamentary Discourse across Time and Space*. John Benjamins.
- Alexander, M., Dallachy, F., Piao, S., Baron, A., & Rayson, P. (2015). Metaphor, popular science, and semantic tagging: Distant reading with the Historical Thesaurus of English. *Digital Scholarship in the Humanities*, 30(s1), i16–i27.
- Alexander, M., & Davies, M. (2015). *The Hansard Corpus 1803–2005*. <https://www.english-corpora.org/hansard/>
- Alexander, M., & Struan, A. (2013). “In countries so unciviliz’d as those?” The language of incivility and the British experience of the world. In M. Farr & X. Guégan (Eds.), *Experiencing Imperialism: Interdisciplinary and Transnational Perspectives on the Colonial and Post-Colonial British* (pp. 232–249). Palgrave Macmillan.
- Archer, D. (2017). Mapping *Hansard* impression management strategies through time

- and space. *Studia Neophilologica*, 89(s1), 5–20.
- Barkawi, T., & Brighton, S. (2013). Brown Britain: Post-colonial politics and grand strategy. *International Affairs*, 89(5), 1109–1123.
- Baron, A., & Rayson, P. (2008). VARD 2: A tool for dealing with spelling variation in historical corpora. In *Proceedings of the Postgraduate Conference in Corpus Linguistics*.
<https://eprints.lancs.ac.uk/id/eprint/41666/1/BaronRaysonAston2008.pdf>
- Bell, D. (2010). John Stuart Mill on colonies. *Political Theory*, 38(1), 34–64.
- BNC Consortium. (2007). British National Corpus (version 3, BNC XML ed.).
<http://www.natcorp.ox.ac.uk/>
- Bourne, H. R. F. (1900). *The Claims of Uncivilized Races: A paper submitted to the International Congress on Colonial Sociology, held in Paris in August, 1900*. In *Foreign and Commonwealth Office Collection*, King's College London.
- Colley, L. (2014). Empires of writing: Britain, America and constitutions, 1776–1848. *Law and History Review*, 32(2), 237–266.
- Edgerton, D. (2018). *The Rise and Fall of the British Nation. A Twentieth Century History*. Penguin.
- Fitzmaurice, A. (2012). Liberalism and empire in nineteenth-century international law. *The American Historical Review*, 117(1), 122–140.
- Foreign and Commonwealth Office. (2006). *Active Diplomacy for a Changing World: The UK's International Priorities*. TSO.
- Hansard at Huddersfield Project. (2022). *Hansard at Huddersfield*. University of Huddersfield. <https://hansard.hud.ac.uk>.
- Hiltunen, T., Räikkönen, J., & Tyrkkö, J. (2020). Investigating colloquialization in the British parliamentary record in the late 19th and early 20th century. *Language Sciences*, 79, 1–16.
- Jeffries, L., & Sanjurjo-González, H. (2019, May 15–17). *Making corpus-based searching accessible for non-expert users: the case of Hansard* [Paper presentation]. XI International Conference on Corpus Linguistics (CILC 2019) conference. Valencia, Spain.
- Kay, C., Alexander, M., Dallachy, F., Roberts, J., Samuels, M., & Wotherspoon, I. (Eds.). (2022). *The Historical Thesaurus of English* (Version 5.0). <https://ht.ac.uk/>.
- Kumar, K. (2012). Greece and Rome in the British Empire: Contrasting role models. *Journal of British Studies*, 51(1), 76–101.
- Leech, G., Garside, R., & Bryant, M. (1994). CLAWS4: The tagging of the British National Corpus. In *Proceedings of the 15th International Conference on Computational Linguistics (COLING 94)* (pp. 622–628).
<https://ucrel.lancs.ac.uk/papers/coling1994paper.pdf>
- McEnery, T., & Baker, H. (2016). *Corpus Linguistics and 17th-Century Prostitution: Computational Linguistics and History*. Bloomsbury.
- Mollin, S. (2007). The *Hansard* hazard: Gauging the accuracy of British parliamentary transcripts. *Corpora*, 2(2), 187–210.
- Morris, J. (2011). How great is Britain? Power, responsibility and Britain's future global role. *The British Journal of Politics & International Relations*, 13(3), 326–347.
- Piao, S., Dallachy, F., Baron, A., Demmen, J., Wattam, S., Durkin, P., McCracken, J., Rayson, P., & Alexander, M. (2017). A time-sensitive historical thesaurus-based

- semantic tagger for deep semantic annotation. *Computer Speech and Language*, 46, 113-35.
- Pietsch, T. (2013). Rethinking the British world. *Journal of British Studies*, 52(2), 441-463.
- Rayson, P., Archer, D., Piao, S., & McEnery, T. (2004). The UCREL semantic analysis system. In *Proceedings of the Workshop on Beyond Named Entity Recognition Semantic Labelling for NLP Tasks in Association with 4th International Conference on Language Resources and Evaluation* (pp. 7-12) LREC 2004.
https://eprints.lancs.ac.uk/id/eprint/1783/1/usas_lrec04ws.pdf
- Seely, B., & Rogers, J. (2019). *Global Britain – A Twenty-first Century Vision*. Henry Jackson Society. <https://henryjacksonsociety.org/publications/global-britain-a-twenty-first-century-vision/>
- Select Committee on Aboriginal Tribes. (1837). *Report of the Parliamentary Select Committee on Aboriginal Tribes (British Settlements)*. Ball, Chambers, Row and Hatchard.
- Skinner, Q. (1969). Meaning and understanding in the history of ideas. *History and Theory*, 8(1), 3-53.
- Slembrouck, S. (1992). The parliamentary Hansard ‘verbatim’ report: The written construction of spoken discourse. *Language and Literature*, 1(2), 101-19.
- Thompson, G., & Hunston, S. (2000). Evaluation: An introduction. In S. Hunston & G. Thompson (Eds.), *Evaluation in Text: Authorial Stance and the Construction of Discourse* (pp. 1-27). Oxford University Press.
- Toye, R. (2014). The rhetorical culture of the House of Commons after 1918. *History*, 99(335), 270-298.
- Tyrkkö, J. (2019). Kinship references in the British parliament, 1800-2005. In U. Lutzky & M. Nevala (Eds.), *Reference and Identity in Public Discourses* (pp. 97-124). John Benjamins.