

The Historical Semantics of Temporal Comparisons Through the Lens of Digital Humanities

Promises and Pitfalls

Michael Götzelmann, Kirill Postoutenko, Olga Sabelfeld, Willibald Steinmetz

1. Introduction¹

Until recently inquiries into the historical semantics of ‘comparison’ have mostly been limited to explicit reflections on the use and misuse, or the pertinence and impertinence, of comparisons in various historical, scientific, or political settings. The preferred object of study has been what one might call ‘comparatism’, i. e., the more or less conscious application and framing of comparison as a tool in all branches of human knowledge, and it is noteworthy that the early modern period has been very much at the center of attention in this kind of research.² From the point of view of historical semantics such an approach would correspond to a strategy that is primarily interested in the changing meanings of particular words or terms that denote, or directly refer to, comparison. Among these, the noun ‘comparison’ itself and its equivalents in other languages are of course to be considered

1 This article has been devised – and revised – jointly by all four authors. More specifically, Kirill Postoutenko has been responsible for section 2, Olga Sabelfeld for section 3, Michael Götzelmann for section 4, and Willibald Steinmetz for the introductory and concluding remarks (sections 1 and 5).

2 See, for example, *Richter, Melvin*, “That vast Tribe of Ideas”. Competing Concepts and Practices of Comparison in the Political and Social Thought of Eighteenth-Century Europe, in: *Archiv für Begriffsgeschichte* 44 (2002), 199–219; *Eggers, Michael*, Vergleichendes Erkennen: Zur Wissenschaftsgeschichte und Epistemologie des Vergleichs und zur Genealogie der Komparatistik, Heidelberg: Universitätsverlag Winter, 2016; *Grafton, Anthony*, Comparisons Compared: A Study in the Early Modern Roots of Cultural History, in: Renaud Gagné/Simon Goldhill/Geoffrey E. R. Lloyd (eds.), *Regimes of Comparatism: Frameworks of Comparison in History, Religion and Anthropology*, Leiden/Boston: Brill, 2019, 18–48.

in the first place, to which may be added their derivatives (verbs, adjectives, adverbs, etc.) and synonyms, as well as explicit denials of comparability. In this article, for reasons of convenience, we speak of the ‘comparison vocabulary’ or of ‘terms denoting comparison’ when we refer to the semantic field of ‘comparison’ in that narrow sense. Strangely enough, given the importance of comparison as a method in the sciences and humanities, and more broadly as a social practice in all walks of life, a traditional history of the concept ‘comparison’ and adjacent concepts in major European languages, let alone non-European languages, is still only in its very first stages.³ It is obvious that the massive increase of digitized source materials in recent years, even if only searchable in a very basic fashion, has made it much easier to carry out empirically based quantitative and qualitative inquiries into the changing conjunctures, meanings and uses of the comparison vocabulary.

However, the historical semantics of comparison as we understand it in this article moves a step further. While it includes the study of the terms expressly denoting the activity of ‘comparison’ and their changing meanings over time, it also takes account of the fact that most comparisons are ‘performed’ in language without making any use of what we call the comparison vocabulary. Everyone would recognize simple sentences such as ‘*x* is like *y*’, or ‘*j* is better than *k*’, or ‘*t* is unlike *t*₁ in this respect, but the same as *t*₁ in all other respects’ as comparisons, yet, in the vast majority of cases, utterances like these are used in a rather inconspicuous manner in the course of ordinary speech. Most of these ‘comparison-performing utterances’, as we will call them here for convenience’s sake, are used in a routinized fashion without the speakers deliberately intending, or overtly communicating, that they are actually about to make a comparison. Besides studying the comparison vocabulary proper, our more ambitious aim in this article is to prepare the field for identifying the ways in which comparisons are routinely performed through speech acts such as the above-mentioned, by making use of the tools of digital semantics. This requires an extension of historical-semantic

3 For a recent attempt with further references to relevant literature see: *Steinmetz, Willibald*, “Vergleich” – eine begriffsgeschichtliche Skizze, in Angelika Epple/Walter Erhart (eds.), *Die Welt beobachten: Praktiken des Vergleichens*, Frankfurt/New York: Campus Verlag, 2015, 85–134; see also: *Steinmetz, Willibald*, Introduction: Concepts and Practices of Comparison in Modern History, in: Willibald Steinmetz (ed.), *The Force of Comparison: A New Perspective on Modern European History and the Contemporary World*, Oxford/New York: Berghahn Books, 2019, 1–32.

study into a field that has traditionally belonged to the domain of rhetorics, or of philosophical and linguistic attempts at systematizing speech acts, in this case comparative speech acts.⁴

In view of this volume's general aim of exploring ways of using digital tools for the study of comparison we assume that comparison-performing speech acts may be classified so as to produce a manageable number of formalized sequences, or patterns, that can be recognized by 'machines', i. e., specifically tagged digital corpora. The task in front of us is a multiple translation work. In order to use digital tools, we need to translate conventional comparative utterances into a series of standardized sequences (section 2), which will then have to be translated into the specific codes used for queries in digitized and previously tagged corpora (section 3), or into self-defined queries that conform to the codes of freely available taggers applicable to self-defined and purposely digitized text corpora (section 4).

In order to narrow down the scope of our explorative analysis, we have chosen to focus on one particular type of comparison only: temporal comparisons, i. e., comparisons between two or more historical points in time, or between one or more objects over the course of time. The multiple translation work involved is still complicated enough but seems reasonable compared to an attempt at translating the totality of all possible comparisons into the codes suited for digital queries. Another limitation is that we have chosen to study one language only, Modern English. The following sections present our collective work in three steps. Section 2 elaborates on the difficulties and possibilities of systematizing comparison-performing utterances in general, and temporal comparisons in particular. These reflections result in a basic typology of sentences performing temporal comparisons – a typology that subsequently helps to formulate digital queries in existing or self-defined text corpora. Section 3 uses an existing digitized and already tagged corpus, the historical corpus of British parliamentary debates from 1803 to 2005 ('Hansard Corpus'), to explore various ways of identifying temporal comparisons in that corpus and of interpreting the search results in quantitative and qualitative terms. Section 4, by contrast, describes the steps that are needed

4 For interesting typological reflections, but also historical cases studies on comparison-performing speech acts see *Mauz, Andreas/Sass, Hartmut von (eds.), Hermeneutik des Vergleichs. Strukturen, Anwendungen und Grenzen komparativer Verfahren*, Würzburg: Königshausen & Neumann, 2011.

when a researcher wants to work on a corpus of his or her own choice, a corpus that has not been previously digitized and tagged, in this case a series of utopias and dystopias ranging from Thomas More (1516) to contemporary examples of that literary genre. The section is in itself designed as a competitive comparison between ‘man’ and ‘machine’, i. e., between traditional and digital research methods in historical semantics.

2. Typologies of comparative utterances: from abstraction to practice

Before going into the details of a computer-based examination of comparative utterances, it is worth explaining – in a top-down fashion – the general logic behind our analytical procedures. As soon as scholars begin approaching comparative processes in earnest, they are confronted with a barrage of quotes from renowned thinkers of all stripes extolling the virtue and ubiquity of comparison.⁵ Indeed, if we accept the standard approach to comparison as a cross-evaluation of several objects (*comparata*) on the basis of a more or less stable combination of criteria (*tertia*), we almost immediately drown in the sea of introspections, observations and empirical measurements undertaken by most of the living beings on our planet most of the time.⁶ Fortunately, the focus of our project is not *comparisons at large*, but rather the *practice of comparison*, understood as an intersubjective activity occurring in the social world.⁷ Such practices offer no direct access to the mental or perceptual activities that inform comparisons, but present their results in communicative acts that are understandable by others. Thus, when Don John in Shakespeare’s *Much Ado About Nothing* (1612) says about his brother,

5 See one of the earliest examples: “in omni ratiocinatione per comparationem tantum veritatem praecise cognoscimus”. Rene Descartes, “Regulae ad directionem ingenii (1619)”. *Descartes, Rene*, Œuvres. T. X, Paris: Cerf, 1908, 439.

6 See, for instance: Cheah, Pheng, The Material World of Comparison, in: *New Literary History* 40 (2009), 524. Mignolo, Walter D., Who Is Comparing What and Why, in: Rita Felski/Susan Stanford Friedman (eds.), *Comparison: Theories, Approaches, Uses*, Baltimore: John Hopkins University Press, 2013, 99.

7 See: Grave, Johannes, Vergleichen als Praxis. Vorüberlegungen zu einer praxistheoretisch orientierten Untersuchung von Vergleichen, in: Angelika Epple/Walter Erhart (eds.), *Die Welt beobachten: Praktiken des Vergleichens*, Frankfurt a. M.: Campus, 2015, 136–137.

the Prince of Aragon: “I had rather be a canker in a hedge, than a rose in his grace”,⁸ we may be unsure about the exact *tertia* lurking in Don John’s (or Shakespeare’s) mind, but the alleged qualitative difference between Aragon’s hypothetical states – **being a canker in a hedge* and **being a rose in his grace* – are expressed clearly and unambiguously for every competent user of a natural language. In social terms, saying *N would rather [do/feel] X than Y* is practicing comparison, and the systemic nature of language allows us to hope that all statements that conform to this syntactic pattern are comparisons. A number of lexemes (such as the adjectives *similar/different*) and morphemes (the suffix *-er* in comparative adjectives *bigger/smaller* etc.) reinforce this impression of regularity: the form that practices of comparison may take in language is therefore not arbitrary but constrained by the rules and conventions of specific languages. Since grammarians and semioticians know the rules of many, if not most, verbal and visual languages, one could hope that practices of comparison in society may ultimately be reduced to a string of formalized patterns which computers might cherry-pick for us in various kinds of data.

Such hopes are not entirely unfounded: particularly among diehard representatives of linguistic structuralism the notion of language as a set of binary oppositions has always been popular. Roman Jakobson used an example from Lewis Carroll’s *Alice in Wonderland* to present this hypothesis to the general audience: “‘Did you say pig or fig?’ said the Cat. ‘I said pig,’ replied Alice.”⁹ Jakobson commented upon this dialogue: “In this peculiar utterance the feline addressee is attempting to recapture a linguistic choice made by his addressor. In the common code of the Cat and Alice, i. e., in spoken English, the difference between a stop and a continuant consonant, other things being equal, may change the meaning of the message”¹⁰ So, if the semantics of natural languages can possibly be exhaustively described by means of privative oppositions, why not try the same thing with comparisons?

8 Shakespeare, William, *Much Ado about Nothing* [1623], Philadelphia, PA: J.P. Lippincott Company, 2001, 53.

9 Carroll, Lewis, *Alice’s Adventures in Wonderland* [1865], Los Angeles: Enhanced Media, 2016, 32.

10 Jakobson, Roman, *The Cardinal Dichotomy of Language*, in: Ruth Nanda Anshen (ed.), *Language: An Enquiry into its Meaning and Function*, Port Washington and London: Kennikat Press, 1971, 157.

At a first glance, such a typology, if rigorously executed, works quite well. Thus, on a basic level, one might distinguish between grading comparisons, equations, differentiations, statements about comparisons, denials of comparability, etc.; and one might then go on to introduce more fine-grained distinctions between various syntactic or semantic patterns that express these and other basic types.¹¹ Indeed, it seems likely that such a classification of comparative utterances could be arranged into a set of neat privative oppositions (*a* vs. $\neg a$) translatable into computer-friendly formal language without much hassle. A case in point is the difference between comparative (*a* is more/less *n* than *b*) and superlatives (*a* is the most/least *n* in the group [*b*, *c*, *d* ... etc.]). Consider the following utterances by Bossuet and Voltaire seeking to grasp the historical destiny of the Jewish people in the light of the schism between Judaism and Christianity:

“The Jews are more demolished than their temple, or city.”¹²

“It is certain that the Jewish nation is the most singular that the world has ever seen.”¹³

While Bossuet, a French conservative bishop and theoretician of history, takes a well-trodden path back to 70 CE, Voltaire, in the vein of Enlightenment, extends the meaning of the alleged ‘Jewish self-alienation’ beyond religion and history. Consequently, his emphasis is on the general ‘otherness’ of Jews, expressed in their unparalleled “singularity”; Bossuet, in contrast, is focused on the inner “abasement” of the Jewish people, which he judges to be more severe than the destruction of the Second Temple or the corresponding fall of Jerusalem. In grammatical terms, the difference between relative (Bossuet) and absolute (Voltaire) grading comparisons is crystal-clear. The first category is articulated by a combination of the adjective *more* (*plus*) placed just

11 For an elaboration on this see: Postoutenko, Kirill, Preliminary typology of comparative utterances: a tree and some binaries (in print). Inevitably, some of the examples and argumentation are borrowed from this article.

12 “Les Juifs sont plus abattus que leur temple et que leur ville” (Bossuet, *Jacques-Bénigne*, Discours sur l’Histoire universelle [1681], Paris: Garnier-Flammarion, 1966, 274.

13 “La nation juive est la plus singulière qui jamais ait été dans le monde” (Voltaire [*François-Marie Arouet*], Des Juifs [1756], in: André Versaille (ed.), Dictionnaire de la pensée de Voltaire par lui-même, Bruxelles: Editions Complexe, 1994, 688.

before the first *comparatum* and the conjunction *than* (*que*) enclosed between the first and the other *comparata*. The second, in its turn, is conveyed by the combination of the same adjective *plus* with the immediately preceding definite article *le* which – together with the absence of *que* – turns the whole thing into a superlative: *plus ... que* → *le plus...* = *more* → *the most*. In some languages, the situation is even more straightforward as there is a special pair of morphemes responsible for the difference between comparative and superlative degrees in adjectives (for example, *-er/-est* in German). So, on the surface of it, there is an impeccable privative opposition between the absolute and the relative grading comparisons. On the one hand, the differentiation itself looks certain as the association of an expression with one semantic category excludes it from the other, and no third kind of grading comparisons is imaginable. On the other hand, the morphological alternation supporting the distinction seems to be so regular and semantically uniform that any differentiation problems appear highly unlikely. So, up to this point, the prospects of formalizing the typology of comparative utterances look quite rosy.

Alas, as soon as we leave the abstract, platonic meta-grammar hovering high in the sky and descend into the midst of social life, the typology of comparative utterances comes under attack from various sides, including (but not limited to) inconsistencies of verbal languages, social customs and political control. Even in the relatively simple and modest domain of grading comparisons there are myriad ways of verbalizing both absolute and relative grading comparisons which seriously compromise the formalizations attempted above. If we were fixed on the verbal meanings only, we had to conclude that Charles Perrault's mentioning of Louis XIV as "the incomparable prince" merely meant that no sensible comparative utterance with the Sun King as a *comparatum* were possible.¹⁴ In fact, Perrault's statement is an absolute grading comparison echoing many similarly constructed references to Gods, royals and lovers.¹⁵ What is more, in many situations one could observe the obliteration of the very difference between comparatives and

14 "Ces ouvrages divins où tout est admirable, // Sont du temps de Louis, ce prince incomparable." Perrault, Charles, *Parallèle des anciens et des modernes en ce qui regarde les arts et les sciences* [1687], München: Eidos, 1964, 10.

15 See: Luhmann, Niklas, *Liebe als Passion. Zur Kodierung der Intimität*, Frankfurt a. M.: Suhrkamp, 1982, 154. Steinmetz, Willibald, *Above/below, better/worse, or simply different? Metamorphoses of Social Comparison, 1600–1900*, in: Willibald Steinmetz (ed.),

superlatives, as the expressions from the one category are used for expressing its opposite. Thus, when the Baptist rebel John Bunyan speaks of Jesus's "desire that the worst of these worst should in the first place come unto him",¹⁶ he allows for a grading of superlatives, which twists their meanings in a curious way: at any rate, the *worst of these worst* presupposes the difference in meaning between the two identical adjectives, with the second *worst* inevitably losing its superlative semantics. All in all, neither the grammar-based formulae of relative and absolute comparisons nor the very contradistinction between them withstand the test of social practice. Needless to say, the same problems plague many other categories of the typology such as, for example, equations.

Still, if numerous exceptions are taken note of and categorized, we may end up with a workable classification of comparative utterances in which inconsistencies, instead of being swept under the carpet, are seen as crucial markers of social, political and cultural influences upon comparative practices.¹⁷ The first step towards this distant but achievable goal would be the ultimate abandonment of logical and grammatical determinism in the typology of comparative utterances in favor of an interactionist standpoint as pioneered by pragmatist philosophy: Whatever is perceived as a comparison by the reader, listener, etc. should work as such, no matter what shape it takes.¹⁸ For a praxis-oriented theory, this sounds like a good – if not the only – way to go about comparisons. In reality, however, such an approach proves to be no less difficult.

Take the seemingly easy case of *equations*, sentences such as *x is like b*. Even within our eminently cooperative group of four authors we failed to achieve consensus on whether individual sentences taken from a large sample of examples preliminarily earmarked as equations would actually qualify as equations or not. Our sample of supposed equations was taken from dif-

The Force of Comparison: A New Perspective on Modern European History and the Contemporary World, New York and Oxford: Berghahn, 2019, 80–112.

16 Bunyan, John, The Jerusalem Sinner Saved [1688], Edinburgh: The Banner of Truth, 2005, 33.

17 For a first attempt, see: K. Postoutenko, Preliminary Typology.

18 It might be fitting here to draw upon the very beginnings of pragmatist semasiology forestalling "the practical turn": "Before I can think you to mean my world, you must affect my world." William James. The Meaning of Truth [1907]. – James, William, Pragmatism and Four Essays from 'The Meaning of Truth', New York: Meridian Books, 1955, 215.

ferent genres (utopias, political speeches, travel journals, etc.) to ensure the ubiquity of speech acts independent of genre-related languages. And yet, we could not agree whether a sentence like 'He [Hitler] gets more and more like a Caesar' should be considered an equation or not.¹⁹ The difference of opinions on whether individual phrases such as these should be considered an equation was so wide that, for the time being, we chose to abandon the pragmatist rationalization of this particular category of comparison (equation).

In a second attempt, we chose *temporal comparisons* as a test case, i. e., sentences that set apart and compare a state 'before' with a state 'after', or a *now* and a *then*, etc. This time, the rate of agreement among our group of four authors was much higher, and eventually, we came up with a tentative praxis-based formalization of temporal comparisons. Essentially, we saw comparative utterances as being produced by a number of recurrent elements sequentially ordered in sentences, or groups thereof. Among the most frequent elements one could name:

- (1) Time references – both deictic (then, now, in old days, before, after N, today, in the future, earlier, etc.) and intrinsic (1.1.2017, after the war, XX ago, ever, some day);
- (2) Terms denoting either changes occurring over time (rise, drop, increase, decrease, no longer, no more, etc.) or their absence (remain, stand still, etc.);
- (3) Conjunctions and temporal adverbs: than, since, ever, then, as, etc.

The one unquestionably productive sequence is [Time reference/its absence] – ... – term of change/difference – ... – conjunction – ... – time reference. An example of such a sequence would be: *In 1970 there were more smokers than today*. In self-comparisons, i. e., comparisons in which not two or more different objects, but different states of the *same* object in different times are evaluated, a standard sequence would be Object N – ... terms of change or its absence – ... – Object N. An example for that kind of temporal comparison would be: *London remains London*.

These short remarks may suffice to show that temporal comparisons are relatively easy to detect, reasonably distinct and well suited to be typecast

19 Shirer, William, *Berlin Diary – The Journal of a Foreign Correspondent, 1934–1941*, New York: Knopf, 1942, 7 (entry for 20 April 1937).

by means of formalized sequences. For these pragmatical reasons we chose temporal comparisons as a starting point for our inquiries into the advantages and disadvantages of digital tools for elucidating the historical semantics of comparisons more generally.

3. Case study I: temporal comparisons in parliamentary interaction

Parliamentary debates are an important source for an investigation of comparisons, and of temporal comparisons in particular, because of parliaments' specific features as interactional environments. First, the interaction in parliaments is characterized by *pro et contra* argumentation,²⁰ and argumentation generally, according to Chaim Perelman, can hardly avoid the use of comparisons "where several objects are considered in order to evaluate them through their relations to each other".²¹ Secondly, parliamentary interaction is based on deliberative rhetoric that aims at decision making. Parliament offers a political arena for the negotiation of opposite points of view in a competitive mode of speaking between different parties or between opposition and government. The participants of a debate rarely accept or reject statements completely, but most of the time modify, question or add arguments. For these reasons, parliamentary debates are particularly suitable for analyzing comparison-performing speech acts as well as explicit rejections or reevaluations of comparisons, for example when speakers identify certain *tertia* or *comparata* as controversial or start questioning established relations between compared units. Another reason why parliamentary debates are an exceptional source for any inquiry into the changing uses of language is, of course, that in many countries they are now available in digitized form over long time spans. Possibly the highest standard of digitization so far has been achieved for the British parliamentary debates of the 19th and 20th centuries assembled in the Hansard Corpus.

20 Palonen, Kari, Concepts and Debates. Rhetorical Perspectives on Conceptual Change, in: Willibald Steinmetz/Michael Freeden/Javier Fernández Sebastián, Conceptual History in the European Space, New York and London: Berghahn, 2017, 101.

21 Perelman, Chaim/Olbrechts-Tyteca, Lucie, The New Rhetoric: A Treatise on Argumentation, trans. John Wilkinson and Purcell Weaver, 2nd ed., Notre Dame and London: University of Notre Dame Press, 1958, 242.

3.1 The Hansard Corpus 1803–2005: 1.6 billion words and some taggers

The Official Reports of British parliamentary debates provided by Hansard date back to the year 1803 and are available in several digitized versions.²² In 2015, the *Hansard Corpus 1803–2005* (HC) was completed and put online. It contains nearly all the speeches held in both houses of the British parliament, Lords and Commons, from 1803 through to 2005.²³ The uniqueness of this corpus lies in the combination of a big amount of data and a semantic annotation of the entire corpus with the *Historical-Thesaurus-based Semantic Tagger* (HTST).²⁴

The HTST provides an annotation of *all* lexical units in a text under certain grammatical and semantic criteria. The semantic classification of the HTST is based on the *Historical Thesaurus of English* (HT), a project that started in the pre-digital age (1964) and was undertaken at the University of Glasgow.²⁵ The developers of the HTST describe the method of their classification as follows:

“The semantic classification is based primarily on a systematic analysis of the content of the Oxford English Dictionary, with other content from additional dictionaries of English. To this end, words are arranged into categories by the

22 Hansard offers two versions of the *Historic Hansard 1803–2005* <https://www.hansard-corpus.org>: The debates in xml format <http://www.hansard-archive.parliament.uk> and the data base behind the *Hansard-Corpus 1803–2005* to read online <https://api.parliament.uk/historic-hansard/index.html>. The *Hansard Online* website presented by the UK parliament offers alongside the historical debates the recent sittings of both chambers, too <https://hansard.parliament.uk/>. Unfortunately, the very new platform *Hansard at Huddersfield* could not be considered for this article <https://hansard.hud.ac.uk/site/index.php> [accessed: 13.03.2019].

23 For the project see: *University of Glasgow*, SAMUELS Project (Semantic Annotation and Mark-Up for Enhancing Lexical Searches), <https://www.gla.ac.uk/schools/critical/research/fundedresearchprojects/samuels/> [13.03.2019].

24 Piao, Scott *et al.*, A time-sensitive historical thesaurus-based semantic tagger for deep semantic annotation, in: *Computer Speech & Language* 46 (2017), 113–135. This article gives further detailed information on the development and structure of the tagger.

25 <https://ht.ac.uk/>. The researchers in that project also developed the *Historical Thesaurus of Old English*, <https://oldenglishtesaurus.arts.gla.ac.uk/> [accessed: 13.03.2019]. Kay, Christian, Diachronic and synchronic thesauruses, in: Philip Durkin (ed.), *The Oxford Handbook of Lexicography*, Oxford: Oxford University Press, 2015, 367–380.

concepts they express, with successive subdivision of these categories delineating ever more precise sub-concepts within a concept."²⁶

An important asset of the HT is that its taxonomy of words is 'time sensitive', i. e., that it takes account of historical changes in the meanings of words. More than other existing thesauri, the HT is therefore capable of eliminating the ambiguities of words that are due to semantic change in a short-term or *longue durée* perspective. The HTST is the first annotation tool that makes use of the time sensitive semantic classification of the HT. The classification of the HT starts at the highest level with three main categories: *the world*, *the mind*, and *society*. Each of these main categories is then subdivided into several further, ever more precise subcategories with lists of individual words assigned to each of the categories and subcategories. Depending on the historical dictionary entries, an individual word may be allocated to several subcategories or even main categories in the hierarchy.

The tagging of the Hansard Corpus (HC) is based on the hierarchy of the HT but assigns particular codes to each of the categories and subcategories. Thus, the category 'Time' (AM) is subdivided into 'Spending Time' (AM:01), 'Duration' (AM:02), 'Particular Time' (AM:03) and so on. These codes can be used to start search queries, either on their own or – what is more interesting – in combination with searches for other categories or particular words or parts of speech. For an unspecific query a large category (e. g., AM – Time) may be more useful than a smaller one (e. g., AM:04:b – a month/calendar month). Thus, the query **{AM:04:b} * war** (month in combination with the word 'war') results in collections of sentences that contain formulations like 'months of war', 'months after war', 'in November the war would have ended by now', and so on.

The semantically tagged Hansard Corpus thus provides the opportunity to search not only for occurrences of single words, but also for syntactic and semantic patterns. Therefore, our rough typology of comparison-performing utterances (see section 2) has to be further refined and adapted for inquiries in the HC. When starting queries on temporal comparisons in the HC we need to go beyond the formalized basic types of syntactic sequences outlined above. For each lexical unit, the precise function and position in the sequence have to be clearly defined before starting a query.

26 S. Piao et al., *Tagger*, 115.

One of the most advantageous features of the HC is its annotation of time references. Searching for time references on a large scale may be a promising avenue to detect patterns of temporal comparison in the corpus. The HC considers as time references not only numerals referring to dates (1952), temporal adverbs (*yesterday*) or nouns (*future*) but also combinations of different parts of speech which constitute a time reference (*at the present time*). While numerals referring to dates and adverbs or nouns referring to time might perhaps be put together “manually” in a search list, it seems impossible that an individual researcher should be able to imagine all possible combinations of time references in advance and assemble them in a search list. By contrast, the HC provides this search option which will be further discussed below.

Apart from time references, another important indicator of temporal comparisons is the assertion that entities have changed across time or that a change has occurred between two points in time. This is usually expressed by using verbs or nouns denoting change, such as *rise*, *decline*, *increase*, etc. Due to the finely graduated semantic annotation of the HC we are capable of searching for those verbs or nouns without concretizing the terms beforehand. Of course, this feature applies not only to terms denoting change but also to many other semantic fields that we might be interested to study.

Fig. 1: Hansard Corpus 1803–2005, Interface

[illegible]

Figure 1 shows the user interface of the HC. It is divided into three parts that display different kinds of information. Whereas the left-hand side contains options to define search queries, the two boxes on the right-hand side display the search results. The box in the upper right part shows relative or absolute frequencies of the words or sequences searched for by time periods (decades, individual years), whereas the box in the lower right part shows lists of individual occurrences of the words or sequences searched for, sorted by the criteria defined on the left-hand side. Several options are available to define queries. The box labeled ‘display’ on the upper left-hand side refers to the way in which search results will appear in the boxes on the right-hand side. Four options are offered: *List*, *Chart*, *KWIC*, and *Compare*. Depending on the type of queries some of the offered four possibilities may be more useful than others. Thus, for someone interested in a diachronic conceptual history, it may be more significant to see a *chart* showing the relative and absolute frequency of the concept over decades than just a *list* with the number of occurrences per decade. *KWIC* (Keyword in Context) provides a traditional concordance view of words or phrases around the sequence. The last option *compare* allows contrasting two different words or sequences under various criteria (to be defined), for example frequency or collocations.

Figure 2 illustrates the query **compar*** limited to the House of Commons with the display option *chart*. The columns show the number of hits and their frequency per decade. Each column (decade) contains another ten columns with detailed frequency of the lemma for every year in that decade. By choosing one of the years, a list of occurrences containing the lemma in that year appears in the box below. Lemmatization of the verb “compare” or the noun “comparison” would work for every query as well, but the short form with an asterisk “compar*” offers the highest scope of possible parts of speech regarding this word (e. g., comparable, comparatively, etc.).

The next section on the upper left-hand side (‘search string’) is content-related, but for basic searches only the first box (*word[s]*) has to be filled. For a distinct collocation the field *collocates* has to be completed with a single word or a part of speech. By clicking on that field but leaving it empty, all the collocations will be displayed and listed whether by frequency or relevance. *Pos list* helps if one is not sure about the abbreviation for a part of speech and wants to include it in the sequence searched for. Finally, *semantic* refers to the semantic taxonomy described above and allows to search for either categories or individual words.

The next section allocates limitations like *decades*, *speaker*, *parties*, *party in power* or *chamber* (Lords or Commons). It is possible to choose only one parameter, several parameters at the same time, or none at all. In the latter case the query will run all through the whole corpus. Unfortunately, a search by subject matter of debates is not possible.

The last two sections ‘sorting and limits’ and ‘options’ improve the usability of the displayed results.

In the HC, it is possible to formulate queries that combine semantical categories, parts of speech (general and specific, e. g., general adjective, general comparative adjective, general superlative adjective, etc.), particular words and expressions, unspecified synonyms of words, and words in all possible grammatical forms. Punctuation can also be included and treated like a lexical unit in a sequence. The following chart shows some important elements of queries for temporal comparisons in the HC.

Table 1: Important elements of queries for temporal comparisons in the HC

Syntax1 ^a	Meaning within the query
*	Wildcard for any number of letters (even in a word, e. g. un*) but only one word in a sentence. Necessary for part of speech categories (see below).
[pos]	Part of speech in an exact grammatical form.
[pos*]	Part of speech in every grammatical form (wildcard).
?	Wildcard for one letter, can be part of a word on every position in this word.
=	Synonyms of a word. If [[=word]] HC searches for this word and its synonyms, whether noun, verb, or adjective, or other.
[]	For all forms of a word.
-	Exclusion of exact words or parts of speech.
	Separation of words or parts of speech to search for any of these words.

a For a more extensive overview see: Alexander, Marc/Davies, Mark, Hansard Corpus 1803–2005, https://www.hansard-corpus.org/help/syntax_e.asp [accessed: 29.04.2019].

Semantic categories 1 ^b			
{AM}	Time		
{AM:04}	Period	{AM:08}	Relative time
{AM:04:a}	Year	{AM:08:b}	The Present (time)
{AM:04:a:01}	Season	{AM:08:c}	The past
{AM:04:b}	A months/calendar month	{AM:08:c:01}	Historical period
{AM:04:c}	A day/twenty-four hours	{AM:08:c:02}	Antecedence/being earlier
		{AM:08:c:03}	Oldness/ancientness
		{AM:08:d}	The future/time to come
		{AM:08:d:02}	Newness/novelty, recency

- b For all semantic categories and options to search for a semantic category of an arbitrarily chosen word see *M. Alexander/M. Davies*, Hansard Corpus 1803–2005, <https://www.hansard-corpus.org/SemTags1.asp> [accessed: 29.04.2019].

Parts of speech 1 ^c			
[a*]	article	[nn*]	common noun, neutral for number
[csa*]	as (as conjunction)	[p*]	Pronoun
[csn*]	than (as conjunction)	[pphs2*]	3rd person plural subjective personal pronoun (they)
[d*]	demonstrative	[ppis2*]	1st person plural subjective personal pronoun (we)
[j*]	general adjective	[rgr*]	comparative degree adverb (more, less)
[jjr*]	general comparative adjective	[v*]	Verb
[jst*]	general superlative adjective	[word*],[pos*]	word as particular part of speech
[n*]	Noun	[[=word]]	word and its synonyms

- c For a list of all included parts of speech that is taken from the UCREL tagger see *M. Alexander/M. Davies*, Hansard Corpus 1803–2005, <http://ucrel.lancs.ac.uk/claws7tags.html> [accessed: 29.04.2019]; Some possible combinations for queries based on parts of speech. *M. Alexander/M. Davies*, Hansard Corpus 1803–2005, https://www.hansard-corpus.org/help/posList_e.asp [accessed: 29.04.2019].

3.2 Six basic sequences to search for temporal comparisons

Having briefly described the HC and its capabilities, we will now introduce six formalized sequences of semantic and syntactic components that may help to detect temporal comparisons in the Hansard Corpus. The list does not claim to be exhaustive but still contains essential components that can be extended and modified for further inquiry. Also, every formalized sequence consists of several varieties of nearly identic sequences that only differ from each other in the space between the units or the specification of semantic components. The example sentences in the following paragraphs are mainly taken from parliamentary debates of the post-war periods after World War I and II.

Our first formalized sequence contains sentences using comparatives. Obviously, the core of those sentences is formed by a general comparative adjective (*better, faster, stronger*) and the conjunction *than*. These components can be combined in different ways, for example: {AM} **than** * {AM}; [jjr*] [csn*] [csa*] {AM}; [n*] [v*] [jir*] [csn*] {AM}. These constructions may also lead to sentences of the type *never happier than before* that are meant as superlatives but formulated as a graded comparison. That case illustrates once more that our classification is based on purely formal criteria, not on considerations related to content. Here is an example sentence, taken from the debate on the Representation of the People Bill of 1948:

“But, as I have said, there was no contract and no bargain made at the Speaker’s Conference to the effect that the present constituencies were going to be perpetuated and that the agreements were going to be perpetuated forever, because the world is moving *quicker today than ever before*, and this Parliament must move with it.”²⁷

The formulation of the sequences depends on one’s own research interests and on assumptions about their probable yield in terms of frequency or relevance. These assumptions, however, may be deceptive, e. g., the sequence {AM} [jir*] * {AM} seems to render a high output, but in fact offers only 1021 sentences in 200 years of House of Commons’ debates. This figure appears

27 Arthur Woodburn (LAB), Presentation of the People Bill [sic!]. HC Deb 17 February 1948 vol 447 c1021.

high, and there are many highly interesting sentences among the 1021 hits, but in terms of the amount of data available in the HC the figure obviously involves only a very small fraction of all temporal comparisons contained in the corpus. This finding illustrates that several modifications and search queries are needed for each formalized sequence to provide a sufficiently high number of sentences.

Our second formalized sequence uses superlatives and similar terms hinting at the highest possible degree of something. With regard to temporal comparisons the words *never* combined with a time reference (**never** * **{AM}**) and *ever* combined with a superlative adjective (**[ijt*]** * **ever**) are the most obvious cases in point. By entering the latter sequence and searching for phrases spoken in the House of Commons displayed as *list*, it is striking to see that superlative phrases of this type are mostly used to highlight presented numerical figures or the high quantitative amount of other entities, whether in a positive or negative way. Here is a typical example sentence taken from a debate on food prices in 1955:

"I am asking if he will state just how far stabilisation and reduction of prices has taken place. Is he not aware of the fact that he has not mentioned that at all? Prices have been going up and up and have reached the *highest level ever* in peace or war."²⁸

Rhetorically speaking, superlative expressions such as these seem not to require any evidence and a particular *comparatum* but are based merely on statements that demand to act on the matter.

Our third formalized sequence contains at its center conjunctions or temporal adverbs that serve as indicators of temporal comparisons. For example, the query as **[j*]** as **{AM}** may render sentences in which this sequence constitutes only the first part of a more complex comparison, as in the following example, taken from a debate on the international situation in September 1944:

"During this war America, Russia and Britain have made common sacrifices. If we trust each other, as we are doing now, why cannot we trust each other when the war is over, for the future of peace and for the good of the world?"

28 Arthur Lewis (LAB), Question on Food Prices. HC Deb 24 November 1955 vol 546 c1638.

Unless we are prepared to do that, and join the common pool, and set up an international police force on the lines I have indicated, then *as sure as night follows day*, there will be another war – perhaps in the next generation.”²⁹

That means that our formalized sequences do not always need to contain the complete comparison, but that it may be sufficient to identify parts of it, or hints at their existence, in the sentence before or after. The same observation applies to the sequence **{AM} * not yet** as well, as in the following example, taken from a statement of David Lloyd George in a debate on a possible better future for Russia beyond Bolshevism in April 1919:

“You cannot carry on a great country upon rude and wild principles such as those which are inculcated by the Bolsheviks. When Bolshevism, as we know it and as Russia to her sorrow has known it, disappears, then the time will come for another effort at re-establishing peace in Russia. But that *time is not yet*. We must have patience, and we must have faith. You are dealing with a nation which has been misgoverned for centuries, and been defeated and trampled to the ground, largely, let us admit, owing to the corruption, the inefficiency, and the treachery of its own governors. Its losses have been colossal. All that largely accounts for the real frenzy that seized upon a great people. That is why a nation which has gone through untold horrors has abandoned itself for the moment to fantastic and hysterical experiments. But there are unmistakable signs that Russia is emerging from the trouble. When that time comes, when she is once more sane, calm, and normal, we shall make peace in Russia. Until we can make peace in Russia, it is idle to say that the world is at peace.”³⁰

Our fourth formalized sequence takes advantage of the HC’s capacity to search synonyms of certain words. As explained above, in the HC it is no longer necessary to create a list with possible synonyms, but there is a function to search directly for all synonyms. As a matter of course, not every synonym helps to identify comparisons. For example, the sequence **{AM} * [[=change]]**

29 John Joseph Tinker (LAB), War and international situation. HC Deb 28 September 1944 vol 403 c549.

30 David Lloyd George (LIB), Question on a Barrier against Bolshevism. HC Deb 16 April 1919 vol 114 c2944-5.

provides a large number of sentences containing the terms *amendment* and *trade* that are in certain contexts used as synonyms of *change*. In these cases there are of course no comparisons. However, this problem is compensated by displaying the found sentences as *list*. The upper right section offers sequences sorted by frequency, and particular expressions can be selected from the list. Further useful expressions of change with time reference are **used to [v*] {AM}** and **than * used to be**.

Our fifth formalized sequence tries to identify temporal comparisons in which terms denoting comparisons are used explicitly or which contain particular words that refer to a relation between entities, words such as ‘superior to’. For that type of query it is sufficient to enter just the relevant word with an asterisk in the box *word(s)* and the abbreviation for time reference **{AM}** in the box *collocates*. Depending on whether equations or differentiating comparisons are in the center of consideration, all or part of the following vocabularies may be important for this task: **analog***, **compar***, **contrast***, **simil***, **inferior***, **superior***, **progress***, etc. Another word that belongs to the semantic field of comparison but needs more elaborate queries is ‘same’. One possibility could be **{AM} * same * {AM}**. For a simple exploration of synonyms of ‘comparison’ HC offers the synonym construction **[[=compare]]** to extend one’s own assumptions regarding possible meanings of related words.

Our sixth formalized sequence tries to collect sentences in which comparisons are contested, modified, or otherwise made a topic of debate. Expressions that may indicate a contestation of comparison or lead to new comparisons substituting old ones are, for example, * **[be] * [[difference] | [discrepance] | [change] | [distinction] | [dissimilarity]]** or * **[be] * analogy**. Here is an example sentence, taken from a debate on the Proportional Representation Bill of 1921:

“At the present time the electors of all parties are glad to put confidence in and to seek help and advice from their Member, because they regard him as ‘Our Member.’ They know who he is. Under the new system it will be absolutely impossible for a candidate at an election to visit every polling district in a huge area, and it will be very difficult, indeed, for him to remain in touch with the whole of the wide and scattered constituency after he has become elected. *There is no analogy or comparison* at all between the case of a Member

of Parliament and that of a guardian elected for a little town in Ireland. I am not a bit impressed by those arguments from Ireland."³¹

The example shows a rejection of comparability, and in this case a rejection that uses a temporal comparison to justify the statement. However, without a close reading of the sentence in context it would have been impossible to identify the temporal components of the comparison. The time references *at the present time* and *new system* are uttered two sentences before the searched sequence and thus beyond the typical collocation span. This makes it nearly impossible to formalize such sequences in advance.

The application of all six formalized sequences works quite well and provides a large amount of comparison-performing speech acts. In later steps of our project the formalized sequences may help to identify comparisons in other text corpora, particularly in those that are dominated by argumentation and persuasion. For a detailed historical semantic analysis the number of sentences resulting from our searches in the HC is far too big to analyze them historically. At this point it is essential to apply limits to the inquiry. Different options are available in the HC to achieve this. One possibility entails bringing the functioning of HC closer to traditional historical research by narrowing the sample by time, speaker, party and other settings. For questions focused on actors, the HC offers not only a limitation by party membership of the speaker but also a party membership related to the governing party. This search function is helpful for examining shifting usages of concepts or patterns of argument within a party over time, between political parties, or between the roles of government and opposition. Another possibility to build a more specific sample is to create a list with relevant speakers and to use the corpus of all their speeches or to select particular speeches of different speakers. For that purpose, the section *decade/speaker* contains the option 'create list' where one can set several parameters – chamber, speaker, time period and party.

As already mentioned, the HC does not allow to select debates by subject matter in order to find out, for example, in which thematic contexts comparisons are most likely to appear. However, our sixth sequence that relates to comparisons that are contested or modified or otherwise made a topic may

31 Gerald Hurst (CON), Proportional Representation Bill. HC Deb 05 April 1921 vol 140 c637.

help to identify these contexts. In addition, our fifth category, referring to the terms denoting comparison (*compar**, etc.) and their collocations, may also be useful for this purpose.

3.3 Between distant and close reading

From a historian's perspective, there are advantages and disadvantages of working with the HC. The most notable advantage is of course the sheer amount of data available in the HC and the web server's capacity to allow fast searches.³² Searches for single words or collocations that may be historically relevant in political contexts can be effectuated on a large scale and with unprecedented speed. For conceptual historians who want to test hypotheses about occurrences of particular words or combinations of words, this is an invaluable tool. Beyond such simple searches, the time required for 'learning' to formulate precise queries and for overcoming disappointments in several sessions of 'trial and error' has to be measured against the possible gains in terms of surprise findings that would be impossible for anyone reading the debates in a traditional, hermeneutical way. Working with the HC and its inbuilt tagger (HTST) thus requires a willingness to switch frequently between the techniques of distant reading and subsequent close reading, or the other way round. Moreover, the lack of a search function by subject matter of debates requires various detours in order to create a more limited text corpus that may be better suited to respond to the historian's particular interests. Two other – very practical – disadvantages are the limitation of access and various restrictions for saving the relevant text materials.

Despite the disadvantages, however, the uniqueness of the HC lies not only in the collection of almost all parliamentary speeches from 200 years but in the multiple annotations of every single lexical unit in the corpus. Therefore, the HC optimally supports 'traditional' conceptual-history searches of single terms and their collocations, and, due to the precise documentation of speakers, parties, political roles, dates and debates, it also provides important information to contextualize the findings; this is particularly helpful for researchers interested in the conceptual horizons of historical actors. Fur-

32 Anthony, Laurence, A critical look at software tools in corpus linguistics, in: *Linguistic Research* 30 (2013), 152–153.

thermore, the HTST allows very elaborated queries in the sense of parts of speech, patterns of sentences, rhetorical figures, punctuation and further criteria. Even if the exceptionality of British parliamentary language has to be kept in mind, the Hansard debates are thus an outstanding source for the history of political language and rhetoric in general.

4. Case study II: man vs. machine – towards a universal tool for identifying comparisons

The literary genre of utopias should be a good source for identifying comparisons, as it inherently deals with differences. Already in Thomas More's eponymous *Utopia* (1516),³³ which deals with the differences between European nations and the imagined state of the island Utopia, comparisons have the function of closing the gaps between the fictional storyline and a potential reader who is situated in his or her own present. In the subgenre of uchronia, which owes its name to Charles Renouvier's utopian novel *Uchronie* (1876),³⁴ the problem of closing gaps in space and time via comparison is even more complex. If the genre as a whole is apparently depending on the use of comparisons in order to function as a literary text, it follows that comparisons, especially temporal comparisons, should appear quite frequently in utopias and uchronias.

The importance of utopias and uchronias should be evident for historians. In general, time is the dimension historians are used to work with; according to Niklas Luhmann this is due to the function of synchronization, which historiography fulfills in a differentiated social system.³⁵ And according to Luhmann, as well as Reinhart Koselleck, time is multidimensional, consisting of pasts, presents and futures, which are interwoven.³⁶ Thus, in the last forty years, the 'futures past' have been a research topic for historians in the

33 More, Thomas, *Utopia*, trans. Gilbert Burnet [1551], New York: Cassel & Co., 1901.

34 Renouvier, Charles, *Uchronie – L'Utopie dans l'Histoire*, Paris: Bureau de la critique philosophique, 1876.

35 See Luhmann, Niklas, *Weltzeit und Systemgeschichte: Über Beziehungen zwischen Zeithorizonten und sozialen Strukturen gesellschaftlicher Systeme*, in: *Soziologie und Sozialgeschichte, Special Issue* 16 (1973), 81–115.

36 Koselleck, Reinhart, *Zeitschichten: Studien zur Historik*, Frankfurt a. M.: Suhrkamp, 2000.

same way as the bygone past, and in that context the literary genres of utopia and uchronia have become a privileged source for historians to access the interplay of time-layers which are so visible in them.

For the purposes of this article we have put together a small text corpus of utopias/uchronias that had to fulfill two basic conditions. First, for pragmatic reasons and better comparability of results, all texts had to be in English, whether in the original or a (contemporary) translation. Secondly, the corpus had to cover a long time span from the sixteenth century to the present. Obviously the first text to consider was Thomas More's *Utopia*. In addition, a small selection of the classics of the genre had to be considered as well, starting with the unfinished manuscript of Francis Bacon's *New Atlantis* (1627) and James Harrington's *The Commonwealth of Oceana* (1656), which are both, like More's *Utopia*, discourses of state philosophy.³⁷ Also added to the corpus was the English translation of Louis-Sebastien Mercier's uchronia *The Year 2500* (1795, the French original *L'an 2440* appeared in 1771).³⁸ Reinhart Koselleck considered Mercier's *L'an 2440* to be the very first uchronia, but this assumption can no longer be maintained since there are at least three earlier texts by Samuel Madden (1733), Heinrich Gottlob Justi (1759) and an anonymous author (1763) which may well be described as uchronias (but these rather obscure texts were not included in the corpus).³⁹ The next example to be included was the first dystopia, the inversion of the positive state or possible future-to-be to the negative: Mary Shelley's *The Last Man* (1825).⁴⁰

37 Bacon, Francis, *The advancement of learning and New Atlantis* [1627], London: Oxford University Press, 1969.

38 Mercier, Louis-Sebastien, *Memoirs of the Year 2500*, trans. Thomas Dobson [1795], Boston: Gregg Press, 1977.

39 See: R. Koselleck, *Zeitschichten*, 131; Madden, Samuel, *Memoirs of the Twentieth Century*, London: Unknown, 1733; Justi, Heinrich Gottlob, *Untersuchung ob etwan die heutigen europäischen Völker Lust haben möchten, dereinst Menschenfresser oder wenigstens Hottentotten zu werden*, Philadelphia: Jacob Heinrich Lowe, 1759; *Anonymus*, *The Reign of Georg VI: 1900–1925*, London: Printed for W. Niccoll at the Paper Mill in St. Paul's Churchyard, 1763, ed. C. Oman (London: Reprinted by Rivingtons, 34, King Street Covent Garden, 1899).

40 Shelley, Mary, *The Last Man* (1825), ed. Hugh Luke, Lincoln: University of Nevada Press, 1965.

Following in chronological order, there are three thematic clusters. The first cluster is formed by novels of a socialistic future, which were popular at the end of the nineteenth century on both sides of the Atlantic. This cluster contains Edward Bellamy's *Looking Backward: 2000–1887* (1888) and Jack London's *The Iron Heel* (1908), which is written like a dystopia, but is actually a utopia. Also part of that cluster are William Morris's *News from Nowhere* (1890) and Eugen Richter's libertarian answer *Pictures of a Socialistic Future* (1893, the German original *Sozialdemokratische Zukunftsbilder* appeared in 1890).⁴¹ Aldous Huxley's *Brave New World* (1932) is included as a stand-alone, because it is neither certain whether this is a utopian or dystopian novel nor does it fit into any of the thematically organized clusters.⁴²

The second cluster, which deals with the rise of communism and fascism, was created around George Orwell's *1984* (1949). It also contains Yevgeny Zamyatin's *WE* (1924), which Orwell used as a template for *1984*, Karin Boye's *Kalloccain* (Swedish original 1940, English translation 1966) and the Post-WWII *Fahrenheit 451* (first published 1953) by Ray Bradbury.⁴³

With Murray Leinster's short story *A Logic Named Joe* (1946) begins the third thematic cluster, which deals with the potentially negative influence of computer technology on human society. The same applies to Isaac Asimov's *I, Robot* (1950), Philip K. Dick's *Do Androids Dream of Electric Sheep?* (1968), William Gibson's *Neuromancer* (1984), and Adam Sternbergh's *Shovel Ready* (2014).⁴⁴ An exception of a kind is Ernest Callenbach's *Ecotopia* (1975), which is the only yet found utopian novel of the second half of the 20th

41 Bellamy, Edward, *Looking Backward: From 2000 to 1887*, 1888, Cambridge, Massachusetts: Harvard University Press, 1978. London, Jack, *The Iron Heel*, 1908, Auckland: Floating Press 2009. Morris, William, *News from Nowhere, or an Epoch of Rest*, being some chapters from *A Utopian Romance*, 1890, New York, Bombay and Calcutta: Longmans, Green & Co. 1908. Richter, Eugene, *Pictures of a Socialistic Future: Freely adapted from Bebel*, London: Swan Sonnenschein, 1893.

42 Huxley, Aldous, *Brave New World* [1932], Middlesex: Penguin Books, 1971.

43 Orwell, George, 1984, London: Secker & Warburg 1949; London: Penguin, 2016. Zamyatin, Eugene, *WE*, trans. Gregory Zilboorg, New York: E. P. Dutton, 1924; New York: E. P. Dutton 1952. Boye, Karin, *Kalloccain*, trans. Gustav Lannestock, Madison, Wisconsin: University of Wisconsin Press, 1966. Bradbury, Ray, *Fahrenheit 451*, Philadelphia: Chelsea House Publ., 2001.

44 Leister, Murray, *A Logic Named Joe*, in: *Astounding Science Fiction* 37 (1946), 139–154. Dick, Philipp K., *Do Androids Dream of Electric Sheep?*, New York: Ballantine, 1968. Asimov, Isaac, *I, Robot*, New York, Doubleday & Co., 1950. Gibson, William, *Neuromancer*,

century that deals with practical solutions to the upcoming ecological crisis.⁴⁵ Anthony Burgess's *A Clockwork Orange* (1962) is also part of the corpus, but will be ignored in this study, due to the used sociolect with its own dictionary.⁴⁶

Once the corpus was put together, there were two possible options to search for (temporal) comparisons. The first option was to simply read the novels, mark all the sentences that include comparisons 'by hand', and transcribe them into a document. The second option was to digitize all texts and apply one or more search tools to identify (temporal) comparisons in the corpus. Since the digitization and adaptation of available tools in collaboration with the INF project proved to be time-consuming, both options were pursued in parallel; therefore this case study has amounted in fact to an experimental comparison between the outcome and relative efficiency of the work of 'man and machine'.

Reading those novels was certainly not 'simple' reading, but it implied the careful questioning and categorizing of every sentence in order to decide whether it actually was (or contained) a comparison or not. This kind of focused reading took a lot of time. The 2151 pages⁴⁷ of the corpus took approximately 143 hours of focused reading, by an average reading speed of 15 pages per hour. It was comforting to know, however, that there would be no way around reading the novels anyway, because the analysis of the functions and uses of comparisons in the literary texts cannot be done without careful scrutiny. Still, the process of marking the sentences and transferring them into digital data was a task which a machine could certainly have done much faster than a human. My estimate is that this process, if done by a machine, would have saved more than 60 % of the invested time because in that case the reading speed could have been increased to 40 pages per hour.

New York: Ace Books, 1984. *Sternbergh, Adam*, Shovel Ready, New York: Random House, 2014.

45 *Callenbach, Ernest*, *Ecotopia: The Notebooks and Reports of William Weston*, Berkeley: Banyan Tree Books 1975.

46 *Burgess, Anthony*, *A Clockwork Orange: The Restored Edition*, London: Random House, 1962; London, Random House, 2012.

47 Based on the printed DinA4-Pages of the .txt-documents.

4.1 Translating for a machine

Before a machine can be employed there is the problem of text format. Most of the novels which were written before 1950 were available on archive.org or Gutenberg.org, normally already in .txt-format. In that case the texts only had to be ‘cleansed’ of headers, rights-disclaimers, introductions of the editors, etc. The other texts had to be scanned to PDF first, then converted into .txt-format, and from there into .xml-format to then be tagged. The tagger used for all documents was USAS, an open source tagger that does lemmatization and semantic annotations. The tagging procedure is necessary to enable the ‘machine’ to search for more than just individual signs, e. g., for inflected adjectives or for words semantically tagged as time references. As a query-processing-software we used the *IMS Corpus Workbench* in the web-based version (CWB).⁴⁸ This query-processing software is a search-engine that allows inserting simple or complex syntax-based patterns and shows the results on the screen.

Working with such a tool demands to formalize syntactic sequences beforehand, in our case sequences that express temporal comparisons. At this stage, it proved to be tremendously helpful that the entire corpus had already been read and the temporal comparisons marked ‘by hand’. Thus, several hundred temporal comparisons had already been identified and classified.

Among these, two main types were particularly salient. The first type may be called inter-timeline-comparison: One or more *comparata* X situated in a time A ($T=0$ or $T=0 \pm x$) are compared to one or more *comparata* Y situated in a time B ($T=0$ or $T=0 \pm x$) with regard to one or more *tertia* V. An example sentence for type I, taken from Ernest Callenbach’s *Ecotopia*, would be:

“In Ecotopia (time A) at that time (time A; indicator $T=T-x$), as in the United States (time B) now (time B; indicator $T=0$), such areas (*comparata*) were

48 The Claws7-tagset and USAS-semantic-tagset were developed at the University Centre for Computer Corpus Research on Language (UCREL) at Lancaster University. The open source query-processing software was also developed at UCREL. For more information see ucrel.lancs.ac.uk [accessed: 01.09.2019].

mainly devoted to (tertium=areal usage) factories, warehouses, sewage plants, railroad yards, dumps, and other unsavory uses.”⁴⁹

The second type may be called inner-timeline-comparison. In the utopian genre this type of comparison is often used when characters in the novel reflect upon themselves and their development across time: One *comparatum* X situated in a present ($T=0$) is compared to the same *comparatum* X at another point on a timeline ($T=T \pm x$) with regard to one or more *tertia* V An example sentence for type II, taken from Eugen Richter’s *Pictures of a Socialistic Future* would be:

“I (comparatum) am sorry to say that I can now ($T=0$) no longer (indicator $T=T-x$) take my meals (tertium=meal times) with my wife except on Sundays, as I have been accustomed for the last (indicator $T=T-x$) twenty five years ($x=25$ years).”⁵⁰

Defining the *comparata* in a search list beforehand is hardly possible; they may be characters in the novel, groups, places, objects, abstract thoughts or historical periods. However, our traditional reading and marking procedure has shown that comparative references to the ancient Greeks or Romans, to “the savage”, “caveman” or “primitive man”, to earlier stages of civilization, to an “other world” or an “old(er) world”, to something “medieval” or “ancient”, appear quite frequently in utopias. Similarly, in inner-timeline comparisons, “boyhood”, “childhood”, “boy” and “girl” are often used as comparative references for the novel character at his or her respective present ($T=0$).

Much easier to define beforehand are the temporal indicators. They comprise temporal prepositions or adverbs like before, last (day, month, year, century, era, seasons etc.), yore, formerly, heretofore, (than) ever, since, earlier, etc., but also possessive pronouns followed by the word “time” (In my time ..., In your time.), as well as specified or unspecified temporal relations (earlier time/that time) and the expression “used to be”. The indicators for the respective present are relatively few, mostly $T=0$ is represented by the word “now”, followed by “no longer” or “no more”. Sometimes, however, a reference

49 E. Callenbach, *Ecotopia*, 87.

50 E. Richter, *Pictures of a Socialistic Future*, 43.

to a present may be hidden in a personal pronoun like “we”, as in a sentence such as “*In the 19th century ... We ...*” or even by simply opposing “*You ... We ...*” without any temporal indicator at all. While someone who reads in the traditional, ‘analogue’ style will often be able to detect the temporal comparison that is hidden in such an opposition, it seems unlikely that a machine can ever be trained to such a degree that it will correctly select these and similarly ‘hidden’ temporal comparisons.

Having established two main types of temporal comparison and described their way of functioning in utopian novels, we may now proceed to formulate queries in the language used by Corpus Workbench. In the following table the first column shows the syntactical query, while the second column explains its respective aims. All queries were constructed with the general or semantic tagset, and they include most of the above mentioned indicators. Not included are references to different stages of civilization, so misses are to be expected.

Table 2: Set of queries with taggers

QNo	Query	Description
Q1	[sem contains "T1.1.1"][*] [sem contains "T1.1.2"] within s;	Searches for a semantically tagged word relating to the past and a semantically tagged word relating to the present within a sentence.
Q2	[sem contains "T1.1.1"][*] [sem contains "T1.1.2"] within 50;	Searches for a semantically tagged word relating to the past and a semantically tagged word relating to the present within a range of 50 words.
Q3	[sem contains "T1.1.1"][*] [sem contains "T1.1.2"] within 75;	Searches for a semantically tagged word relating to the past and a semantically tagged word relating to the present within a range of 75 words.
Q4	[sem contains "T1.1.2"][*] [pos="RRR"][*] [sem contains "T1.1.1"] within s;	Searches for a semantically tagged word relating to the past, followed by comparative general adverb and a semantically tagged word relating to the present within a sentence.
Q5	[sem contains "T1.1.2"][*] [pos="RRR"][*] [sem contains "T1.1.1"] within 50;	Searches for a semantically tagged word relating to the past, followed by comparative general adverb and a semantically tagged word relating to the present within a range of 50 words.

QNo	Query	Description
Q6	[sem contains "T1.1.2"][*] [pos="RRR"][*][sem contains "T1.1.1"] within 75;	Searches for a semantically tagged word relating to the past, followed by comparative general adverb and a semantically tagged word relating to the present within a range of 75 words.
Q7	[sem contains "T1.1"][*][sem contains "A6"] within s;	Searches for a semantically tagged word relating to the past and a general comparative term within a sentence.
Q8	[sem contains "T1.1"][*][sem contains "A6"] within 50;	Searches for a semantically tagged word relating to the past and a general comparative term within a range of 50 words.
Q9	[sem contains "T1.1"][*][sem contains "A6"] within 75;	Searches for a semantically tagged word relating to the past and a general comparative term within a range of 75 words.
Q10	[pos="RRR"][*][pos="NNT1"] [pos="NNT2"] within s;	Searches for a comparative adverb, followed by a temporal noun, in singular or plural, within a sentence.
Q11	[pos="RRR"][*][pos="NNT1"] [pos="NNT2"] within 50;	Searches for a comparative adverb, followed by a temporal noun, in singular or plural, within a range of 50 words.
Q12	[pos="RRR"][*][pos="NNT1"] [pos="NNT2"] within 75;	Searches for a comparative adverb, followed by a temporal noun, in singular or plural, within a range of 75 words.

4.2 Counting for a machine

Once the search queries had been formulated they could be put to a test. To make things a little bit easier, we selected only five utopian novels: Thomas More's *Utopia* as the oldest, Adam Sternbergh's *Shovel Ready* as the most recent one, and in between one for each of the thematic clusters: Jack London's *The Iron Heel* representing the socialism-cluster, George Orwell's *1984* representing the fascism/communism-cluster and Philip K. Dick's *Do Androids Dream of Electric Sheep?* for the fear-of-the-machine-cluster. For a first attempt, this selection seemed sufficiently large to represent different variants of English and of temporal comparison.

The first test served to determine the success rate of the machine in identifying those comparisons that had already been found 'by hand'. The

following table shows the total number of temporal comparisons identified ‘by hand’ in the column next to the book title, followed by the result for each query, the total number for all queries, and in the last column the success rate for the machine in percent.

Table 3: Comparison of results man vs. machine

Title	Total by hand	Q1	Q2	Q3	Q4	Q5	Q7	Q8	Q9	Q10	Q11	Q12	Total machine	%
Utopia	19	0	0	0	0	0	0	0	0	0	0	0	0	0
Iron Heel	71	1	0	5	0	0	0	0	0	8	0	0	14	19.72
1984	50	2	0	12	1	0	0	0	0	9	0	0	22	44
Ecotopia	35	2	3	0	0	0	0	0	0	0	0	0	5	14.29
Shovel Ready	39	0	0	3	0	0	0	0	0	0	0	0	3	7.69
Total	214	5	3	20	1	0	0	0	0	17	0	0	44	20.56

The results are extremely disappointing for our digital tools. They only found 44 temporal comparisons out of the 214 found ‘by hand’. This amounts to a very low overall success rate of 20.56 percent, which seems to indicate that the software is not worth giving it further tries. It worked still reasonably well for *1984* but had very poor results for *Utopia* and *Shovel Ready*. Why there were no hits at all for *Utopia* is hard to understand. It may either be that there was a problem in the pipeline, i. e., in the process of formatting and tagging, or that the maximum search range of 75 words is too small because many of the temporal comparisons are made within a longer sentence or across several long sentences. The main reason why there were so few hits in *Shovel Ready* could be the high context dependence of temporal comparisons within the story and the short and often incomplete sentences that are used. The story of *Shovel Ready* builds on a terror act with a dirty bomb which alters everyday life in New York City in the near future. Spademan, the protagonist, a former garbageman and now a contract killer, lost his wife in the attack. In the limnopsphere, a more progressive form of the internet, he encounters a digital version of his wife again. “My wife. In the same dress I last saw her in.” The reader knows that his wife is dead and that the scene takes place after the attack. The machine should have recognized this as well because the semantic

tagger T.1.1.1 should have identified ‘last’ as a word relating to the past and ‘same’ should have been identified by the tagset category A6 as part of a comparison. So the fault does not seem to be a result of the novel’s specific language, but rather indicates a problem with the tagset, which does not work as accurately as hoped for. But one should stress that it did not work as badly as the table indicates. For the above table does *not* include those comparisons that the machine found *in addition* to those found by manual search. These are now shown in table 4:

Table 4: Additional temporal comparisons found by the machine

Title	Total
Utopia	0
Iron Heel	15
1984	29
Ecotopia	13
Shovel Ready	1
Total	58

These results show that neither man nor machine are perfect. There are misses on both sides. Apparently, the machine is unable to ‘understand’ the story which is created by the text as a whole and will therefore miss temporal comparisons that are obvious for human readers. On the other hand, it may help us to compensate for gaps in concentration during the process of reading. The tagsets which come along with Corpus Workbench are far from perfect.

Better results might be achieved in the future by coding one’s own semantic tagsets. One could do this either by editing the existing Corpus Workbench tagset. Or – and that is another test we have made – by creating searches that use our own above mentioned indicators. Admittedly, this is a much less elegant solution, but it turned out that it worked a little bit better than the tagset. On such a basis an archetype for a different kind of semantic tagger could be created. Table 5 shows the set of queries used for this test.

Table 5: Set of queries with our own indicators

QNo	Query	Description
Q13	"before" "past" "last" "century" "yore" "formerly" "heretofore" ("than" [* "ever"]) "caveman" "savage" "era" "your" "time" "since" "earlier" "no" "more" "no" "longer" "primitive" "man" "now" "ancient" "roman" "greek" "to-day" "today" "old" "years" "month" "day" "never" "modern" "history" "first" "once" "present" within s;	Searches for a set of indicators within a sentence.
Q14	"before" "past" "last" "century" "yore" "formerly" "heretofore" ("than" [* "ever"]) "caveman" "savage" "era" "your" "time" "since" "earlier" "no" "more" "no" "longer" "primitive" "man" "now" "ancient" "roman" "greek" "to-day" "today" "old" "years" "month" "day" "never" "modern" "history" "first" "once" "present" within 50;	Searches for a set of indicators within a range of 50 words.
Q15	"before" "past" "last" "century" "yore" "formerly" "heretofore" ("than" [* "ever"]) "caveman" "savage" "era" "your" "time" "since" "earlier" "no" "more" "no" "longer" "primitive" "man" "now" "ancient" "roman" "greek" "to-day" "today" "old" "years" "month" "day" "never" "modern" "history" "first" "once" "present" within 75;	Searches for a set of indicators within a range of 75 words.

Again, these queries were applied to the same five utopias mentioned above. This time, the column next to the book title contains the total of all comparisons found manually and those that the machine found in addition. The results are as follows.

Table 6: Comparing results of man vs. machine with our own indicators

Title	Total	Q13	Q14	Q15	Total	%
Utopia	19	0	0	0	0	0
Iron Heel	86	64	2	0	66	76.74
1984	79	32	4	0	36	45.57
Ecotopia	38	26	0	0	26	68.2
Shovel Ready	40	6	1	0	7	17.0
Total	262	128	7	0	135	51.53

Searching only for our self-defined indicators of temporal comparisons led to much better results than the queries based on the tagsets. With the exception of *Utopia*, which still did not show any valid result, the overall success rate of 51.53 percent is, as expected, much higher than that produced by the application of the tagsets. One might argue that this method is a kind of self-delusion considering that we are searching for something that is already known to be there. The next step might be to apply our self-defined search queries to a different text corpus and compare the results with those achieved by a traditional reading taking place after that experiment.

4.3 And the winner is ...

Overall there is much work left to do. At this stage of our cooperative project, we are still very far away from our desired aim of disposing of a 'machine' that can reliably identify temporal comparisons in the English language. We will either have to modify the Corpus Workbench's tagsets or construct our own tagger for indicators of temporal comparisons. Using a syntax which only relies on the indicators would be a third and not so elegant way to solve this problem, but even this would still require a revision of the indicators and a test with another corpus.

5. Conclusion

Our case studies on the already digitized and tagged Hansard Corpus on the one hand and on the self-defined corpus of utopias on the other hand have shown that the successful identification of temporal comparisons in texts by means of digital tools is highly dependent on the quality of the tagsets used. Whereas the search facilities provided by the HC interface, supported by the HTST tagger, have rendered some valuable results, the rate of 'hits' produced by a freely available tagger (Corpus Workbench) applied to the corpus of utopias was rather low, especially if put into relation to the amount of time needed to prepare the textual material and to adapt the codes used by the tagset to our specific questions. Even if the problems discussed in this article can be solved in a satisfactory way, it may still be a long way to go from the reliable identification of certain types of comparison-performing utterances to a historical analysis of their changing use patterns in certain historical

contexts or literary genres. However, there is no way of denying the progress in terms of acceleration brought by searchable corpora and elaborated taggers like the HTST, compared to the efforts and time that would have been necessary in the pre-digital age to effectuate a traditional conceptual history of 'comparison', its adjacent concepts, and basic forms of articulating comparisons. The result of our explorative inquiry, therefore, is in itself a temporal comparison, and one that is advantageous for the present compared to a not too remote past.

Bibliography

- Alexander, Marc/Davies, Mark*, Hansard Corpus 1803–2005, https://www.hansard-corpus.org/help/syntax_e.asp [accessed: 29.04.2019].
- Anonymus*, The Reign of Georg VI: 1900–1925, London: Printed for W. Niccoll at the Paper Mill in St. Paul's Churchyard, 1763.
- Anthony, Laurence*, A critical look at software tools in corpus linguistics, in: *Linguistic Research* 30 (2013), 152–153.
- Asimov, Isaac*, I, Robot, New York, Doubleday & Co., 1950.
- Bacon, Francis*, The advancement of learning and New Atlantis [1627], London: Oxford University Press, 1969.
- Bellamy, Edward*, Looking Backward: From 2000 to 1887, 1888, Cambridge, Massachusetts: Harvard University Press, 1978.
- Bossuet, Jacques-Bénigne*, Discours sur l'Histoire universelle [1681], Paris: Garnier-Flammarion, 1966.
- Boye, Karin*, Kalloccain, trans. Gustav Lannestock, Madison, Wisconsin: University of Wisconsin Press, 1966.
- Bradburry, Ray*, Fahrenheit 451, Philadelphia: Chelsea House Publ., 2001.
- Bunyan, John*, The Jerusalem Sinner Saved [1688], Edinburgh: The Banner of Truth, 2005.
- Burgess, Anthony*, A Clockwork Orange: The Restored Edition, London: Random House, 1962; London, Random House, 2012.
- Callenbach, Ernest*, Ecotopia: The Notebooks and Reports of William Weston, Berkeley: Banyan Tree Books 1975.
- Carroll, Lewis*, Alice's Adventures in Wonderland [1865], Los Angeles: Enhanced Media, 2016.

- Cheah, Pheng*, The Material World of Comparison, in: *New Literary History* 40 (2009).
- Descartes, Rene*, Œuvres. T. X, Paris: Cerf, 1908.
- Dick, Philipp K.*, *Do Androids Dream of Electric Sheep?*, New York: Ballantine, 1968.
- Eggers, Michael*, *Vergleichendes Erkennen: Zur Wissenschaftsgeschichte und Epistemologie des Vergleichs und zur Genealogie der Komparatistik*, Heidelberg: Universitätsverlag Winter, 2016.
- Gibson, William*, *Neuromancer*, New York: Ace Books, 1984.
- Grafton, Anthony*, Comparisons Compared: A Study in the Early Modern Roots of Cultural History, in: Renaud Gagné/Simon Goldhill/Geoffrey E. R. Lloyd (eds.), *Regimes of Comparatism: Frameworks of Comparison in History, Religion and Anthropology*, Leiden/Boston: Brill, 2019, 18–48.
- Grave, Johannes*, Vergleichen als Praxis. Vorüberlegungen zu einer praxis-theoretisch orientierten Untersuchung von Vergleichen, in: Angelika Eppe/Walter Erhart (eds.), *Die Welt beobachten: Praktiken des Vergleichens*, Frankfurt a. M.: Campus, 2015, 136–137.
- Huxley, Aldous*, *Brave New World* [1932], Middlesex: Penguin Books, 1971.
- Jakobson, Roman*, The Cardinal Dichotomy of Language, in: Ruth Nanda Anshen (ed.), *Language: An Enquiry into its Meaning and Function*, Port Washington and London: Kennikat Press, 1971.
- James, William*, *Pragmatism and Four Essays from 'The Meaning of Truth'*, New York: Meridian Books, 1955.
- Justi, Heinrich Gottlob*, *Untersuchung ob etwan die heutigen europäischen Völker Lust haben möchten, dereinst Menschenfresser oder wenigstens Hottentotten zu werden*, Philadelphia: Jacob Heinrich Lowe, 1759.
- Kay, Christian*, Diachronic and synchronic thesauruses, in: Philip Durkin (ed.), *The Oxford Handbook of Lexicography*, Oxford: Oxford University Press, 2015, 367–380.
- Koselleck, Reinhart*, *Zeitschichten: Studien zur Historik*, Frankfurt a. M.: Suhrkamp, 2000.
- Leister, Murray*, A Logic Named Joe, in: *Astounding Science Fiction* 37 (1946), 139–154.
- London, Jack*, *The Iron Heel*, 1908, Auckland: Floating Press 2009.
- Luhmann, Niklas*, *Liebe als Passion. Zur Kodierung der Intimität*, Frankfurt a. M.: Suhrkamp, 1982.

- Luhmann, Niklas*, Weltzeit und Systemgeschichte: Über Beziehungen zwischen Zeithorizonten und sozialen Strukturen gesellschaftlicher Systeme, in: *Soziologie und Sozialgeschichte, Special Issue* 16 (1973), 81–115.
- Madden, Samuel*, *Memoirs of the Twentieth Century*, London: Unknown, 1733.
- Mauz, Andreas/Sass, Hartmut von (eds.)*, *Hermeneutik des Vergleichs. Strukturen, Anwendungen und Grenzen komparativer Verfahren*, Würzburg: Königshausen & Neumann, 2011.
- Mercier, Louis-Sebastien*, *Memoirs of the Year 2500*, trans. Thomas Dobson [1795], Boston: Gregg Press, 1977.
- Mignolo, Walter D.*, Who Is Comparing What and Why, in: Rita Felski/Susan Stanford Friedman (eds.), *Comparison: Theories, Approaches, Uses*, Baltimore: John Hopkins University Press, 2013.
- More, Thomas*, *Utopia*, trans. Gilbert Burnet [1551], New York: Cassel & Co., 1901.
- Morris, William*, *News from Nowhere, or an Epoch of Rest, being some chapters from A Utopian Romance*, 1890, New York, Bombay and Calcutta: Longmans, Green & Co. 1908.
- Orwell, George*, 1984, London: Secker & Warburg 1949; London: Penguin, 2016.
- Palonen, Kari*, Concepts and Debates. Rhetorical Perspectives on Conceptual Change, in: Willibald Steinmetz/Michael Freeden/Javier Fernández Sebastián, *Conceptual History in the European Space*, New York and London: Berghahn, 2017, 96–117.
- Perelman, Chaim/Olbrechts-Tyteca, Lucie*, *The New Rhetoric: A Treatise on Argumentation*, trans. John Wilkinson and Purcell Weaver, 2nd ed., Notre Dame and London: University of Notre Dame Press, 1958.
- Perrault, Charles*, *Parallèle des anciens et des modernes en ce qui regarde les arts et les sciences* [1687], München: Eidos, 1964.
- Piao, Scott et al.*, A time-sensitive historical thesaurus-based semantic tagger for deep semantic annotation, in: *Computer Speech & Language* 46 (2017), 113–135.
- Postoutenko, Kirill*, Preliminary typology of comparative utterances: a tree and some binaries (in print).
- Renouvier, Charles*, *Uchronie – L'Utopie dans l'Histoire*, Paris: Bureau de la critique philosophique, 1876.
- Richter, Eugene*, *Pictures of a Socialistic Future: Freely adapted from Bebel*, London: Swan Sonnenschein, 1893.

- Richter, Melvin*, "That vast Tribe of Ideas". Competing Concepts and Practices of Comparison in the Political and Social Thought of Eighteenth-Century Europe, in: *Archiv für Begriffsgeschichte* 44 (2002), 199–219.
- Shelley, Mary*, *The Last Man* (1825), ed. Hugh Luke, Lincoln: University of Nevada Press, 1965.
- Shirer, William*, *Berlin Diary – The Journal of a Foreign Correspondent, 1934–1941*, New York: Knopf, 1942.
- Shakespeare, William*, *Much Ado about Nothing* [1623], Philadelphia, PA: J. P. Lippincott Company, 2001.
- Sternbergh, Adam*, *Shovel Ready*, New York: Random House, 2014.
- Steinmetz, Willibald*, Introduction: Concepts and Practices of Comparison in Modern History, in: Willibald Steinmetz (ed.), *The Force of Comparison: A New Perspective on Modern European History and the Contemporary World*, Oxford/New York: Berghahn Books, 2019, 1–32.
- Steinmetz, Willibald*, Above/below, better/worse, or simply different? Metamorphoses of Social Comparison, 1600–1900, in: Willibald Steinmetz (ed.), *The Force of Comparison: A New Perspective on Modern European History and the Contemporary World*, New York and Oxford: Berghahn, 2019, 80–112.
- Steinmetz, Willibald*, "Vergleich" – eine begriffsgeschichtliche Skizze, in: Angelika Epple/Walter Erhart (eds.), *Die Welt beobachten: Praktiken des Vergleichens*, Frankfurt/New York: Campus Verlag, 2015, 85–134.
- University of Glasgow*, SAMUELS Project (Semantic Annotation and Mark-Up for Enhancing Lexical Searches), <https://www.gla.ac.uk/schools/critical/research/fundedresearchprojects/samuels/> [13.03.2019].
- Voltaire [François-Marie Arouet]*, *Des Juifs* [1756], in: André Versaille (ed.), *Dictionnaire de la pensée de Voltaire par lui-même*, Bruxelles: Editions Complexe, 1994.
- Zamiatin, Eugene*, WE, trans. Gregory Zilboorg, New York: E. P. Dutton, 1924; New York: E. P. Dutton 1952.

