**Distributional Concept Analysis:**

**A Computational model for parsing conceptual forms**

**1. Introduction**

 The journal to which the following is a contribution has, since its inception, been commendably open to different approaches to the history of concepts, and since its first volume in 2005 it has significantly contributed to the heterogenous development of the field across national and disciplinary boundaries. As it moves into its second decade, calls for a ‘transdisciplinary history of concepts’ and for ‘comparative transnational’[[1]](#footnote-1) approaches feature more prominently in its pages, and this might be taken to imply that some of the earlier head-scratching around the methodological grounding of the field is no longer appropriate, or even very common. In spite of this trend the following paper goes back to some grounding questions in order to provide an example of how some experimental methodologies which adapt computational and statistical techniques developed in both information extraction and computational linguistics might contribute to the heterogeneity of the field.

 Perhaps the most basic question we mean to invoke is this: are concepts the same thing as words? We suspect that many readers of this journal will find such a question naïve or inconsequential, as Willibald Steinmetz does in his 2012 contribution when he indicates in a note that ‘I leave out the tedious discussion about the concept of concept’, preferring instead to work with the self-declared imprecision of the term ‘concept’:

 Thus I speak of concept throughout the article although, strictly speaking, it would be correct to distinguish carefully between the terms (or words) referring to concepts and the concepts themselves.[[2]](#footnote-2)

Our purpose here is not to criticize a significant contribution to the ongoing reception of Reinhart Koselleck’s work, but to underline the awkward and consistent fact that most research into conceptual history, in so far as it has settled into a discipline (or inter-discipline) with clear outlines, assumes what is most commonly taken to be the case, namely that there is a difference between a label for a concept – say the word ‘justice’ in English – and what Steinmetz calls the ‘concept itself’. Intuitively this seems right and it prompts observations of the following kind, taken from a review of a book by Brent Nongbri, *Before Religion: A History of a Modern Concept* written by Helge Arsheim. The author notes that tracking the history of the concept of religion from the ancient world to our own time ‘ultimately comes down to the somewhat basic and blunt observation that the ancient world had no word or corresponding concept to our modern-day conception of religion’.[[3]](#footnote-3) The trip in the road here is ‘no word *or* corresponding concept’. Hedging whether it would be enough to identify the lack of *a term* for religion, Arsheim does what we all do: invoke the assumption that words and concepts are not the same thing and leave us hanging as to whether the ‘or’ marks a real or hard distinction between word and concept as opposed to a soft or inconsequential one. This hedging is common in the literature.[[4]](#footnote-4)

 The following contribution to the on-going development of the field of conceptual history emerges from a four-year project (name withheld for reviewing purposes) which took as its starting point the axiom that words and concepts are not the same kind of thing: they have different ontologies.[[5]](#footnote-5) If one starts here, one immediately has to consider how a putative history of concepts is to be distinguished from a history of the meanings of words.[[6]](#footnote-6) The following sketches out a possible answer that is intended to be exploratory.

 Unsurprisingly, this issue has been addressed in many of the contributions to this journal since its inception and we find that what has emerged is a (weak) consensus around the following headline proposal:

 Conceptual history entails assessing semantic changes and exploring the diverse historical settings where concepts are semantically recast, while comparative conceptual history maps those transnational paths.[[7]](#footnote-7)

Although there is surely much to be learned from tracking the history of semantic changes (both across natural languages and within them) we find that a question is here being begged: does this assume an identity between types of things, in our terms a shared ontology between words and concepts? No one would argue with the fact that ‘semantic change’ must be based in the altering uses of words – Reinhart Koselleck, one of the principal guiding figures for some versions of the history of concepts, explicitly recognises this in his introduction to the *Geschichtliche Grundbegriffe*: ‘…this work [the *Geschichtliche Grundbegriffe* ] explores how modern and old words begin to overlap and shift their meanings’.[[8]](#footnote-8) One might observe that saying words have histories of use, and therefore meaning, does not preclude the possibility that concepts may also alter over time. But the rub here is alter how? If we say – or more commonly assume – that the *meanings* of concepts change over time, in what sense are these meanings to be distinguished from the meanings of words? [[9]](#footnote-9) If they are not so distinguished why make a case for a *history of concepts* as opposed to semantic history in general?[[10]](#footnote-10)

 Of course, Koselleck himself famously insisted that concepts exceed words: ‘The concept is bound to a word, but is at the same time more than a word…’[[11]](#footnote-11) and the way in which he and his followers seek to further this observation is to insist on what might be called parallel semantic fields, one holding for words and the other for concepts. This allows one to mount the hypothesis that the meanings of words and concepts may sometimes overlap or converge and at others be discrete or distant from each other. So, to return to our opening axiom, that words and concepts have distinctive ontologies, we now need to ask if the *meanings* of each are ontologically distinct as well, which, to be absolutely clear, is to ask if the meaning of a word is *the same kind of thing* as the meaning of a concept.

 Some work in the history of concepts addresses this question by expanding the field within which concepts are assumed to operate. Such an expanded field was always central to Koselleck’s project since he was not solely interested in plotting the history of the changing meaning of words or concepts, but in the social and political uses to which such concepts were put.[[12]](#footnote-12) For others, such as the historian of epistemology Hans-Jorg Rheinberger, this expanded field must include not only words but also deeds and things.[[13]](#footnote-13) But even when the discursive in its largest sense is moved to the centre of attention a residual assumption persists, namely that the study of concepts must always entail inquiry into their meaning howsoever that meaning may be constructed, or what kind of thing it may be. The history of concepts as it has emerged is, therefore, at base a semantically motivated field of enquiry.[[14]](#footnote-14) Although there are very good reasons for this investment in the diachronic study of meaning we believe that it hampers attempts to model concepts as distinct entities from words. Difficult as this is, the following sets out some ways in which one might begin to dampen down our interest in semantics as the primary motivation for studying concepts diachronically so as to open up for inspection something that heretofore has been impossible to see: the functions by which concepts operate.[[15]](#footnote-15) In the following paper, which is intended to be exploratory and not definitive, we sketch some ways in which a putative conceptual functionality may be identified from the distributions of lexis in large datasets of language use. We acknowledge that our method, based as it is upon lexical co-occurrence data, does not decisively break the tie between words and concepts because as Jan Ifverson notes: ‘Since concepts are expressed in words, this means that they are always tied to words’.[[16]](#footnote-16) We also see great merit in an approach that identifies larger units of lexical operation, in the literature referred to as semantic fields, as more amenable to specifically conceptual rather than word history. But what if these larger units of lexical operation are not only semantically motivated but also determined by an underlying set of rules of formation and operation that have as yet unexamined – perhaps tangential - connection to meaning as such?[[17]](#footnote-17) These objects of study – semantic fields or networks – are constructions and for the most part conceptual historians have based them upon meaning relations. Now, however, we have the ability to gather information on large constellated lexical behaviour through computational methods which do not necessarily assume a model of semantic equivalence, even if some computational linguists use these methods to discover the basis for such equivalence.[[18]](#footnote-18) Our hypothesis, then, is that we may be able to inspect conceptual function at a level that is pre-semantic or even perhaps non-semantic. If such functions change over time we may then be able to identify one of the *causes* for change in the meanings of concepts. The project from which the paper emerges has been testing such a hypothesis and our method has been first to establish functional descriptions and analyses of concepts based on data derived from word distributions in massive datasets of printed text.[[19]](#footnote-19) Our second step, to be explored in a subsequent paper, is to investigate such data driven descriptions diachronically and then to map any identified changes onto diachronic investigations of changing conceptual meanings.[[20]](#footnote-20)

**2. Background**

We begin with some background on the computational methods we have adapted from disciplines that rarely connect to the history of concepts: cognitive science and computational linguistics. Such methods are also commonly employed in research into natural language processing where projects are directed at tasks such as machine translation, analogy solving, question-answering, and natural language inference systems. We note that considerable success in these domains has been achieved by researchers using distributional semantic models derived from large text corpora.[[21]](#footnote-21)

Irrespective of theoretical motivations, the computational implementations of these methods have much in common. Word, phrase, or document meanings are approximated by deciding on a word-distance window within which to count word co-occurrences or compare the paradigmatic context, and counts of word or context co-occurrence are tabulated into a vector. The vectors can be compared directly to measure word associations, or combined into a matrix to measure the similarity of documents.

One of the most widely applied document-based methods today is a generative probabilistic model of the relationship between words and documents known as topic modelling.[[22]](#footnote-22) Topic modelling discovers groups of associated words by modelling a process in which documents are generated by selecting from subsets of words with varying probabilities. The resulting clusters of associated words are intended to capture topics in the corpus, and to allow comparison of documents by topic and measurement of diachronic change in topic emphasis. The method has been widely applied outside of computer science, for example to measure political attention and literary style.[[23]](#footnote-23)

Applications that focus on measuring word rather than document association often use paradigmatic similarity of contexts shorter than document length.[[24]](#footnote-24) The meaning of a word in such models is represented as a point embedded in a high-dimensional space defined by the contexts in which its occurrences are counted. The context may be a short window of words around the target word or defined by a grammatical relation discovered by automatic syntactic parsing of the text.[[25]](#footnote-25) Since the number of possible contexts may be many times the size of the vocabulary, dimension-reduction methods such as singular value decomposition are often used to reduce the complexity of the model, with the drawback that the resulting dimensions are no longer directly interpretable. Word-context models with dimensionality reduction result in a “word vector”, usually of the order of a few hundred dimensions --- these are commonly referred to as “word embedding” models. An efficient and widely used implementation of this kind of model is the Word2Vec package, which uses a neural network trained to predict word contexts, and encodes the meaning of the word in its parameters rather than explicitly counting word contexts.[[26]](#footnote-26)

Some work in computational linguistics describes types of word co-occurrence in terms of Saussure’s syntagmatic and paradigmatic relations: words that tend to occur close to one another in a passage of text have a syntagmatic relation, while words that tend to have similar contexts - one word may be easily substituted for the other - have a paradigmatic relation.[[27]](#footnote-27) These different kinds of contextual similarity reveal different conceptual relations: syntagmatic similarity captures meronymy, phrasal association, and general conceptual association, while paradigmatic similarity captures synonymy, antonymy, and hyponymy.

An advantage of syntagmatic counts is that the type of semantic association required can be precisely defined. Fine-grained relations such as that between ‘dog’ and ‘bark’ are more clearly instantiated in syntagmatic co-occurrence patterns. Researchers have used specific co-occurrence patterns to learn precise kinds of semantic relation.[[28]](#footnote-28) For example, a meronymic relation between word X and word Y may be determined by computational analysis of counts for phrases like “*Xs consisting of Y*” in a large corpus, whereas syntagmatic patterns for arbitrary semantic relations can be learned from example seed pairs.[[29]](#footnote-29) Descriptions of distributional semantic methods focus on measuring and evaluating word similarity, but although it is not always explicitly stated, the possibility that these models encode information that could be considered conceptual rather than simply lexical is recognised in the literature.

**3.1 Advantages and disadvantages of existing methods for detecting conceptual coherence**

There are numerous well-established approaches in computational linguistics for identifying statistical associations among words on the basis of their patterns of co-occurrence in text corpora.[[30]](#footnote-30) However, a measure that would serve as a foundation for discovering conceptual structure within large textual datasets needs to satisfy additional desiderata. We have therefore developed a measure to satisfy three particularly important requirements for our purposes, namely *transparency, sensitivity to frequency*, and *sensitivity to distance.[[31]](#footnote-31)*

The reasoning behind our methodology can be summed up as follows:

1. *Transparency*: An ideal measure of association should not be a “black box.” If such a measure identifies two words in a corpus as having a strong association to each other, it should be possible to understand why.
2. *Sensitivity to frequency and data sparsity*: A word (e.g. “democracy”) will likely co-occur more frequently with words that are very frequent in the corpus as a whole (“was”, “of”) than with less frequent words (“government”, “aristocracy”), even though “democracy” and “government” serve similar functions. An ideal measure of association should therefore take frequency into account in a way that also considers *data sparsity*. That is to say, no matter how large one’s corpus is, it will always contain a large number of words whose counts are relatively infrequent and less statistically reliable as a result. This difference in statistical reliability should somehow be accounted for as well.
3. *Sensitivity to distance*: Because we wish to capture associations that lay outside narrowly semantic operations, it is important that our measure be able to distinguish how strongly associated two words are at one distance (e.g., 100 words away) from another (e.g., 5 words away).

The standard procedures for deciding whether two words *x* and *y* ‘co-occur’ entail determining (a) whether *y* appears in a “window” of text that extends some specific number of words to the left and/or right of *x*, or (b) whether *x* and *y* appear in the same document. The most common way to implement (a) is for the window of text to begin at *x* and extend outward in both directions*.* As such, a classic “window” of 100 words will include all words that appear just 1, 2, 3… words away, which is not desirable if we wish to exclude words that primarily only appear with *x* in adjectival phrases, idiomatic expressions, and other relations that have more to do with syntax than conceptual relatedness. One reason we use the term ‘co-association’ rather than ‘co-occurrence’ for these “long-range” co-occurrences is that the phrase “co-occurrence” frequently implies that the two words appear in the same phrase or collocation. Another reason is its distinctness from “association,” which is not necessarily symmetric.[[32]](#footnote-32) In contrast, if A co-associates with B the reverse must necessarily be true as well. We believe that such standard procedures do not meet all of the desiderata described above. Neural networks, topic models, and word embedding models all produce famously non-transparent mathematical representations, although there has been some research aimed at improving their interpretability.[[33]](#footnote-33)

Simple measures of co-occurrence are much more transparent. For example, *pointwise mutual information* is frequently used either on its own as a measure of lexical association, or as a starting point for the construction of more complex measures.[[34]](#footnote-34) PMI is given as

|  |  |
| --- | --- |
|  | (1) |

where P(*x*) and P(*y*) can each be approximated as the frequency of *x* and *y* (respectively) divided by the total number of words in the corpus, and P(*x, y*) as the number of times that x and y co-occur divided by the total number of words in the corpus.[[35]](#footnote-35) P(x,y)/P(x)P(y) has an intuitive interpretation: it simply expresses the ratio between the number of times x and y co-occur, divided by the number of times one would expect them to appear if their appearances throughout the corpus were randomly distributed. A ratio of 2 would indicate that one is twice as likely to see *x* and *y* together than one would expect to if these words were randomly and uniformly distributed throughout the corpus. That is, if we assume that *x* and *y* were to retain the same corpus-wide frequencies, but their positions relative to other words in the corpus were completely random. What qualifies as a ‘co-occurrence’ is up to the user of the measure and the most popular choice is appearance in the same window or document.

PMI is highly transparent—two words will rank as highly associated if and only if they appear together in the same contexts more frequently than would be expected by chance, and the contexts in which they appear together in the corpus can be directly inspected. Although PMI is sensitive to the global frequency with which words appear in a corpus, it is insensitive to *data sparsity* – the fact that the most infrequent words in the corpus also provide the least reliable information. If *x* and *y* each appear 100 times in a corpus, and co-occur with each other on 99 of those instances, this should be treated as more important than if x and y only appear once in a corpus and happen to co-occur in that singular instance. However, PMI does not penalize rare, statistically unreliable events in any way. As a result, if one were to rank the words in a corpus that co-occurred with some fixed word *x* according to their PMI with *x*, one would find an inverse relationship between rank and frequency: the higher the rank on the list, the lower the frequency.[[36]](#footnote-36) A common band-aid is to impose a minimum frequency or co-occurrence count, an approach that mitigates the problem but does not remove it.[[37]](#footnote-37) In sum, most simple measures of co-occurrence tend to be transparent but have varying degrees of success with respect to their approach to frequency and data sparsity[[38]](#footnote-38). Approaches such as topic models and neural networks tend to be less transparent.

Finally, although most practitioners define co-occurrence in terms of appearance in the same window or document, this is more a matter of convention than necessity. Window-based measures can be readily adapted to ‘distance-based’ measures: one can specify that in order to count as a co-association with *x*, a word *y* must appear some distance *d* words away from *x,* plus or minus a word or two. This is the approach we have taken because we wish to capture data on patterns of lexical distribution that moves from close up where semantic or syntactic coherence is strongest to far away where it is weakest. By taking this wider view we mean to discover relations of binding that go beyond or underpin strictly local semantic ties. Thus we use what could be called a ‘sliding window,’ as it involves a window of a fixed size that, rather than being centered on *x* itself, is centered on a word *d* words away. The investigations presented here employ a sliding window of five words (e.g., *y* will co-associate with *x* at distance *d* if it appears anywhere between *d – 2* and *d + 2* words away from *x*). Other window sizes would have been possible; these particular selections are somewhat arbitrary and we have experimented with a number of different values for “smoothing” the sliding windows in our visualisations.

**3.2 Constructing a measure for the discovery of conceptual coherence**

We have suggested that an ideal measure would combine the transparency of PMI with the ability of more complex methods, such as word embeddings, to exhibit sensitivity to lexical frequency and data sparsity. One possible method is suggested by Levy, Goldberg, and Dagan who reported an intriguing connection between PMI and one of the most popular word embedding models, word2vec SGNS (‘skip-grams with negative sampling’). They note that one of the features of word2veccorresponds to a ‘smoothed’ version of PMI in which “all context counts are raised to the power of α” (pp. 215), and apply this modification, along with many others, to a relatively simple and transparent model of lexical association.[[39]](#footnote-39) They report that among all modifications applied, this was one of the most effective in increasing performance on a range of benchmark tasks, reporting their performance at α = 0.75. They did not hypothesize why this value worked well, except to point out that similar work had found success with this value.

One possible reason for the success of this value is that it may hit the sweet spot in the trade-off between frequency and data sparsity, which are naturally in opposition. An alpha of 1 corresponds to standard, unsmoothed PMI, which ignores data sparsity and gives too much weight to the specious evidence provided by highly infrequent words. However, an alpha of 0 ignores a word’s corpus-wide frequency. We therefore have adopted a simple variant of PMI, which drops the logarithm and introduces a smoothing exponent in the denominator because we are primarily interested in the rank order of words as they are co-associated with a focal word *x*. Since the logarithm does not affect this rank ordering it can be dropped without loss of generality. Doing so highlights the fact that our measure is essentially a small modification to the measure sometimes referred to simply as ‘observed over expected’: the number of times two words are observed in conjunction, divided by the number of times one would expect to see them together by chance. This measure, which we refer to as a *distributional probability factor (dpf)* since it expresses the extent to which a word is predicted to be distributed across the dataset*,* is as follows:

|  |  |
| --- | --- |
| $$dpf\left(x, y\right)= \frac{P(x, y)}{(P\left(x\right)P\left(y\right))^{α}}$$ | (2) |

When we calculated the value of α that eliminated the inverse correlation between rank and frequency described in the previous section, we found the optimal value for our corpus to be 0.78, very close to 0.75. Higher values corresponded to a negative correlation between rank and frequency, while lower values corresponded to a positive correlation. In other words, with an appropriate value of alpha, *dpf* is sensitive to data sparsity as well as frequency, and retains the simplicity and transparency of PMI. Subsequent references in this work to the *dpf* between two words therefore refer to formula (2) with an alpha of 0.78.

 Recall that we estimate *P(x,y)* in this formula by counting ‘co-associations’ – cases in which *y* appears some particular distance *d* away from *x.* By combining this distance-based approach with *dpf,* we have constructed a measure that meets all three of our fundamental desiderata. The next sections describe how this measure is employed to build a ‘profile’ of a lexical co-association with respect to a particular time period and distance, and how these profiles can be used in our investigation of concepts.

**3.3 Constructing a profile of lexical co-associations**

Computing *dpf* scores between all possible pairs of words in a corpus (between a focal and bound word) yields an enormous range of values. Thousands of pairs appear high to the eye—but how high is high? In this kind of environment, mainstays of parametric statistics, such as the use of *p*-values to determine statistical significance, are not appropriate.[[40]](#footnote-40) How should we decide whether a co-association between two words is strong enough to merit our attention?

The approach we selected is similar to the common technique of determining which components of a factor analysis to preserve by looking for the ‘elbow’ or ‘bend in the curve’ in a scree plot.[[41]](#footnote-41) Consider the list created when one computes the *dpf* between a particular word of interest (the “focal word”) and a large number of other words (Fig. 1). If these *dpf* values are plotted against their rank on the list, one typically obtains a smooth curve (Fig. 2) that is well-fit by a power function. By fitting a power function to this curve and solving for where the derivative is equal to -1, the *dpf* at which the ‘bend in the curve’ occurs can be identified. This value serves as a threshold that separates the ‘short head’ of the curve from the ‘long tail’. In the sections that follow, we refer to words occurring above the threshold as “bound” to the focal word, and their specific *dpf* values as “strength of binding.”

 *Insert table 1*

|  |
| --- |
|  |
| ***Bound word*** | ***dpf*** |
| despotism | 14978.4 |
| despotic | 3928.8 |
| monarchy | 3572.3 |
| arbitrary | 3012.2 |
| discontinuance | 2947.1 |
| governments | 2726.7 |
| aristocracy | 2557.7 |
| decaying | 2548.0 |
| statesmen | 2389.1 |
| revolution | 2291.6 |
| repelling | 2236.0 |
| tyranny | 2222.7 |

 *Insert figure 1*



*Dpf* is similar to PMI and many other measures commonly employed by computational linguists to quantify lexical association, in that the specific identities or lengths of the documents in which the lexical co-associations occur do not figure in the calculation. Like most computational linguists, we see this as an advantage, rather than a disadvantage: the measure is sensitive to the number of co-associations, and not to the number of documents or contexts in which co-associations appear (which would require semi-arbitrary decisions to be made about what counts as a 'document' or 'context'). We should also note that we make no claim as to the specific importance of a distance or span: we investigate *dpf*s at large distances not to suggest that there is anything deeply meaningful about a peak in *dpf* at some specific distance, but rather to distinguish between co-associations that are more likely to be derived from syntax or the 'syntax-semantics interface' from those where such derivation is likely to be weak. If binding is strong in such cases (at long spans) we must posit a reason for this – hence our interest in this as a possible index to a distinctive conceptual ontology.

**4.0 Towards conceptual functionality**

 In this section we introduce a second computational tool that we call the grapher, which supplements the *dpf* ranker described above. The grapher also calculates the relative probability of the co-association of a focal word with one or more bound words but it reports the binding co-efficient (*dpf*) at *every* separate lexical distance in a given range: where the ranker offers a rapid survey of the whole environment in which a focal word is embedded (at a given distance), the grapher enables us to trace in more specific detail the force that the same word exerts on a finite number of specified bound words by calculating the relative probability of a bound word occurring.[[42]](#footnote-42) This relative probability is plotted on the x-axis while the y-axis plots each separate lexical distance within a specified range (in this case, two hundred words before and after the word). The grapher therefore presents a more fine-grained pattern of co-association across large lexical distances; this precision allows us to identify what we call *preferential order distribution*, where a bound word appears more frequently either before or after the focal word.

 Fig. 2 charts the co-association of ‘freedom’ (the focal word) and ‘liberty’. Here we find a pronounced spike close to ‘0’, ie within five words either preceding or following the focal word, but also note that while the tendency for pronounced co-association declines as we move further out, it does not get close to reverting to random distribution (as indicated on the graph by the black line above the baseline). Thus, at a distance of one hundred words ‘liberty’ is likely to occur both before and after freedom almost seven times more likely than in random distribution. When we inspect the close up distance, however, we find a clear preference: ‘liberty’ is more likely to occur immediately before than immediately after ‘freedom’. Is this preference in the eighteenth-century dataset indicative of a habit of thinking as well as a habit of writing for the period?

 *Insert figure 2*



 If we consider for a moment the semantic values of these two words we might say that ‘liberty’ prepares the ground for, produces, or enables ‘freedom’, where this latter term is not to be taken as a pure synonym, but as some kind of adjunct or inflection. When we move from the data returns above to the individual texts that cumulatively produce them, we see precisely such inflections. ‘By this means our liberty becomes a noble freedom’, states Edmund Burke in his *Reflections on the Revolution in France*.[[43]](#footnote-43) Alternatively, such a process may fail to take place, in which case the two terms become desynonymised: ‘This whole state of commercial servitude and civil liberty taken together is certainly not perfect freedom’, Burke similarly states elsewhere, in a 1774 speech on American taxation.[[44]](#footnote-44) We posit that such semantic effects can be separated from the underlying distributional pattern, the *preferred order distribution*, which in this case functions as a kind of ‘seeding’ whereby one term prepares for its bound partner without prejudice to its semantic colouration (which is to say the function of preferred order can ‘seed’ for both synonymic and antonymic freighting). This can be thought of as a conceptual analogue to what the linguist Michael Hoey calls ‘lexical priming’.[[45]](#footnote-45)

 Let us take a slightly more complex example. Fig 3. charts the co-association of two words: ‘sensibility’ (the focal word) and ‘irritability’. Here we can compare the previous example of ‘liberty’ and ‘freedom’ where the co-association is around seven times more frequent than in a random distribution at a distance of one hundred words before the focal word. In the case of ‘irritability’ and ‘sensibility’ that frequency is one hundred times above random. The co-association is still more striking at close distance: we have cropped Fig. 3 so as to make it legible, but the peak on the x-axis indicates a frequency 4,200 times more than what would be predicted in a random distribution.

 *Insert figure 3*



In this case there is no evidence of preferential order distribution: the relative frequency of the bound word is comparable both ante- and post-focal word. Here we can identify a different underlying pattern and function by noting the extent to which both words co-associate frequently not only with one another, but also with a wide range of further common lexical material. In order to capture this we have collated successive searches using the ranker, so as to reveal which bound words appear prominently at the lexical distances of 5, 10, 20, 30 and 100 words away. In collating bound terms across these different distances we surely capture close up syntactic and grammatical ties between words as well as looser ties that can in part be understood in terms of a common topic. But for our purposes of discerning putative functional attributes to concepts we are less interested in the semantic values of individual words or the larger semantic field they comprise than in the data profile of such common sets of lexis. Once again, we are asking if we can discern an underlying pattern to lexical distributions and their accompanying binding profiles which may point towards structures or functions that are not entirely determined by meaning. Table 4 charts those bound words with irritability across the eighteenth century (note that the lists for 1761-1780 and 1781-1800 have been cut off; both continue for another 53 and 78 words respectively) subdivided into specific historical periods (1700–1720, 1721–1740, 1741–1760, 1761–1780 and 1781–1800).

 *Insert table 2*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 1701-1720 | 1721-1740 | 1741-1760 | 1761-1780 | 1781-1800 |
| irritability | 0 | 0 | animal | accumulation | abates |
|  |  |  | animals | acrid | accumulated |
|  |  |  | dr | aliment | accumulation |
|  |  |  | fibres | animal | aion |
|  |  |  | intestines | animals | animal |
|  |  |  | motion | arteries | animals |
|  |  |  | muscles | bark | applications |
|  |  |  | muscular | bladder | arteries |
|  |  |  | nerve | blisters | bark |
|  |  |  | nerves | brain | bladder |
|  |  |  | property | causes | blisters |
|  |  |  | sensibility | ceases | blood |
|  |  |  | system | chronic | bowels |
|  |  |  |  | circulation | brain |
|  |  |  |  | contraction | causes |
|  |  |  |  | convulsions | chronic |
|  |  |  |  | cough | circulation |
|  |  |  |  | debility | constitutions |
|  |  |  |  | degree | contraction |
|  |  |  |  | delirium | convulsions |
|  |  |  |  | Disease | cutaneous |

The data for sensibility has a significantly different tenor (in this case the lists are complete):

Table 3.

 *Insert table 3*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| sensibility | 0 | delicate | feeling | delicacy | affection |
|  |  |  | irritability | delicate | feeling |
|  |  |  | sensibility | emotion | feelings |
|  |  |  |  | feeling | her |
|  |  |  |  | impressions | impressions |
|  |  |  |  | irritability | irritability |
|  |  |  |  | irritation | morbid |
|  |  |  |  | muscles | nerves |
|  |  |  |  | nerves | nervous |
|  |  |  |  | nervous | pathetic |
|  |  |  |  | organs | sensation |
|  |  |  |  | sensation | sensibility |
|  |  |  |  | sensations | sympathy |
|  |  |  |  | sympathy | temperament |
|  |  |  |  | temperament | tenderness |

Such a protocol enables us to identify which focal words consistently bind a large amount of lexical material, and to track this binding profile diachronically. At the start of the century neither has a common bound set across distance; by the century’s close, however, ‘sensibility’ has a common bound set of fifteen words across each of the specified distances. The data for ‘irritability’, by contrast, indicates a still more pronounced transformation: having no common bound set up to 1740, by the final date-slice 1781-1800 we find a set of ninety-nine bound words. Once again we note that this certainly provides evidence for a semantic history but we wish to look away from the meanings of these words and their co-associates in order to ask if we can discern a particular way of behaving, a functionality, that sits as it were under the level of semantics.

 If we look to tables 2 and 3 above we see that by 1741–1760 ‘sensibility’ has begun consistently to bind only three words (variants on itself and irritability), whereas ‘irritability’ binds a number of terms that comprise a discourse of physiology - nerves, muscles, organs, and so forth. If one looks ahead to 1761–1780 one can see that ‘sensibility’ has accrued much of the same lexical material that had constellated around ‘irritability’ much earlier—by this date both words circulate in a recognisable discourse of affect that includes ‘sympathy’, ‘emotion’, and ‘feeling’. One way of understanding these shifts in semantic fields is to consider the wider contexts within which this language operated which would include, for example, a historical account of the development of psychology as a scientific discipline. We believe that this data could certainly contribute to current conceptual histories which raise such research questions. But we mean to propose something more radical: what if the coalescence of such lexical units – call them the bundles or networks of words that build concepts – are not only determined by sense, meaning or topic but also by habits of cultural thinking, conceptualization, based on a small suite of functional operations? If we consider concepts as tools to do something with – as Wittgenstein did – what functions do these tools have, and do they change over time?

**5.0 Towards a history of conceptual functioning**

In this last section we outline some ways in which our experimental methodology might contribute to an enriched history of concepts that tracks both their changing meanings and their altering functionalities. The following figures are derived from the same calculations we have made to discern the *dpf* of terms in our dataset (ECCO). In this case, however, we have exported the data into a network environment which helps us to identify complex patterns of similarity and connection between data points, here represented as words in our dataset. The figures are based on screenshots from these network representations displayed in an interactive web app. These are ‘neighbourhood’ or ‘ego’ graphs of order two; that is, they show the graph containing nodes within at most two steps from a specific focal node or nodes entered by the user. An edge exists between two nodes if their *dpf* association is above a user-specified value. It is also possible to threshold the edge list by association rank.

 The network is drawn using the threejs R package, with a force-directed layout algorithm which models the network mechanically as repelling particles connected by a spring.[[46]](#footnote-46) The result is that in a graph of suitable density and degree, nodes are spaced apart enough to be distinguished, but the edges pull together nodes into clusters that share many relations. Community structure is detected with an unsupervised modularity optimisation algorithm and indicated by colour.  For visual clarity, nodes with only a single connection (degree one nodes) are not displayed.

 As we can see from Table 2 above the common set of bound words across distance for ‘irritability’ increases in size as the century lengthens and the semantic freighting of the growing network – a semantic field – is clear enough: a language of nerves, fibres organs and so forth. When we plot our data in the format of a network graph we should find precisely this displayed, as indeed we do. But we believe that the following figures also provide historical evidence for an alteration in function whereby ‘irritability’ and its associated network of closely bound terms grows in density, generating sub-networks within its close orbit and, *at the same time*, holds off at a distance another network within which ‘sensibility’ operates. Here we find compelling evidence for the alteration in composition, size and operation of semantic fields, for sure, but we also believe these plots to provide a window onto how these entities – note here that the underlying data is derived from co-association patterns at a distance of one hundred words away from a focal word – change their *modus operandi* over time.

 As we would expect from the data in tables 2 and 3 the two terms ‘irritability’ and ‘sensibility’ should have no overlap or connection up through the 1740s:

Fig 4



During the following decade we begin to observe a network becoming established – here we can see that the visualization software draws a thick line between ‘irritability’ and ‘sensibility’ which corresponds to a high degree of relation (that is high *dpf* value):

Fig 5



In the next decade this strong connection between the two terms decreases as ‘fibre’ comes to make the link or ‘bridges’ the two networks. It is also noteworthy that an independent network around ‘nerve(s)’ and ‘sensation’ becomes established with the strength of connection between ‘fibre’ and ‘nerve’ indicated by the thick line.

Fig 6:



By the last decade of the century something very eye-catching happens: the network within which ‘irritability’ operates becomes denser and more complex as ‘sensibility’ correspondingly gets simpler, and the line of connection between the two terms becomes etiolated and weaker:

Fig 7



It is unwise to read these network plots as directly representing changes in kind of relation – essentially reading them as substantives rather than as conveniences according with the underlying mathematics which takes complex multi-dimensional vectors and fits them to a visually legible form. Notwithstanding this, the structure of these plots is based upon invariant data: they are telling us *something*. Here the elongation of the distance between the node ‘irritability’ and ‘sensibility’ corresponds to a weakening of *dpf* values and, we believe, indicates that a particular force of repulsion or distancing is operating. We might think of irritability and its immediate network of terms as one concept shaping, moving and accommodating itself to an adjacent concept, that represented by sensibility and its immediate network of terms. And the relations between the networks as mediated by specific terms that function as bridges, connectors or junctions enabling movement between regions in the overall network. The task of precisely identifying and distinguishing such functions lies ahead, but we hope the direction of travel is clear.

**6. Conclusion**

This paper is speculative and intended to be exploratory. The construction of the computational entities we have presented has been based upon data derived from lexical distributions in a massive dataset of printed materials from the Anglophone eighteenth century in order to test the hypothesis that something we call ‘conceptual function’ becomes inspectable. We note that the diachronic metadata of the dataset has not been exploited to the full in our examples, but in three further papers we have explored how our methodology can address more standard questions in the history of concepts. The aim of this paper is to make a theoretical intervention within the field of conceptual history so as to turn exclusive attention away from the meanings of concepts towards a more heterogenous project that includes inquiry into their function and structure.

1. See Julian Bauer, “From “Organisms to World Society”: Steps toward a Conceptual History of Systems Theory, 1880-1980”, *Contributions to the History of Concepts*, 9, no. 2 (Winter 2014): 51-72:54; José María Rosales, “Liberalism’s Historical Diversity: A Comparative Conceptual Exploration”, *Contributions to the History of Concepts*, 8, no. 2 (Winter 2013): 67-82:73 [↑](#footnote-ref-1)
2. Willibald Steinmetz, “Some Thoughts on a History of Twentieth Century German Basic Concepts”, *Contributions to the History of Concepts*, 7, no. 2 (Winter 2012): 87-100:88. [↑](#footnote-ref-2)
3. Helge Arsheim, *Contributions to the History of Concepts* , 9, no. 2 (Winter 2014): 90. [↑](#footnote-ref-3)
4. It can also be identified in the use of ‘and’ in the phrase ‘word(s) and concept(s)’ in which it is not clear if the ‘and’ makes an elision between the two or upholds a hard distinction. See, for example, J.G.A Pocock: “The historian of discourse therefore, operating within this guideline, may review certain key words and concepts in the political vocabulary of the Founders…’, “The American founding in early Modern Perspective”, in *Conceptual Change and the Constitution*, Terence Ball and J.G.A. Pocock, eds. (Lawrence, Kansas: University Press of Kansas, 1988), 55-77, here 56; James Farr, “Throughout I will speak of concepts, not of words, unless the distinction need to be drawn. The distinction frequently need not be drawn because our language is so richly developed that most of our concepts – especially our political concepts – express themselves with matching words which name then explicitly and uniquely.”, “Understanding conceptual change politically,” in *Political Innovation and Conceptual Change*, Terence Ball, James Farr and Russell L. Hanson, eds. (Cambridge: Cambridge University Press, 1989), 24-49, here 27 n. 2; [↑](#footnote-ref-4)
5. The word ‘ontology’ is used in a variety of disciplines and discourses and this richness of use can impede understanding. We use the term here to refer to something quite basic, the difference in category or kind of thing. One might note, for example, that apples do not share the same ontology as motion pictures. We note that in computation and information science the word is used in a technical sense to refer to an artefact designed for a specific purpose ‘which is to enable the modelling of knowledge about some domain’. See <http://tomgruber.org/writing/ontology-definition-2007.htm>. Although we do not mean to invoke this strictly technical sense we find the word helps us focus our own inquiry into how concepts model knowledge in ways that are different to how words do. Noting this does not preclude the possibility that the two entities, words and concepts, might share properties, one of which is meaning. This is commented upon below. [↑](#footnote-ref-5)
6. Contributors to this journal have not been slow to note that Quentin Skinner, for example, is in print indicating that he does not think a history of concepts possible – though he later recast this view. See, among other places, Quentin Skinner: “My almost paradoxical contention is that the various transformations we can hope to chart will not strictly speaking be changes in concepts at all. They will be transformations in the applications of the terms by which our concepts are expressed.”, “Retrospect: Studying Rhetoric and Conceptual Change”, in *Visions of Politics: Regarding Method*, vol. 1 (Cambridge: Cambridge University Press, 2002), 179; Jan-Werner Müller, “On Conceptual History”, in *Rethinking Modern European Intellectual History*, Darrin M. McMahon and Samuel Moyn, eds. (Oxford: Oxford University Press, 2014), 74-93, here 75. [↑](#footnote-ref-6)
7. Jose Maria Rosales, ‘Liberalism’s Historical Diversity: A Comparative Conceptual Exploration’, *Contributions to the History of Concepts*, 8, no. 2 (Winter, 2013): 67-82:73. [↑](#footnote-ref-7)
8. See Reinhart Koselleck, Introduction to GG, *Contributions to the History of Concepts*, 6, no. 1 (Summer, 2011): 1-37:8. For one of many descriptions of Koselleck’s method in the same vein see Jan Ifverson: ‘There is no doubt that for Koselleck, doing conceptual history entailed a word history, or rather a historical semantics, based on the study of language in the sources (*Quellensprache*).’ See *Contributions to the History of Concepts*, 6, no. 1 (Winter 2011): 65-88:72. [↑](#footnote-ref-8)
9. One can see the issue clearly in this contribution to the journal from O. Hidalgo: ‘conceptual history (here understood as the description and analysis of concrete historical semantics, origins, derivation and alterations of concepts)…’, *Contributions to the History of Concepts*, 4 (2008): 176-201:177. How is ‘concrete historical semantics’ – a history of the meanings of words – distinguished from ‘alterations of concepts’? [↑](#footnote-ref-9)
10. The distinction between a history of ideas and a history of concepts – claimed by Elias Jose Palti to be the major difference between Ernst Cassirer’s project and Koselleck’s – does not settle this issue. It may well be that concepts have the outline Koselleck claims for them (the semantic shifts that occur from alterations in the context of use *are inscribed in the concept*) but the ontology of meaning itself remains unchanged. See Elias Jose Palti, “Reinhart Koselleck: His Concept of the Concept and Neo-Kantianism”, *Contributions to the History of Concepts*, 6, no. 2 (Winter, 2011): 1-20: 4 [↑](#footnote-ref-10)
11. Reinhart Koselleck, “Begriffsgeschichte and Social History”, in *Futures Past: On the Semantics of Historical Times* (Cambridge, Mass.: MIT Press, 1985), 84. [↑](#footnote-ref-11)
12. Here we agree with the view that Skinner’s attention to ‘the uses to which words were put’ at different times in history and by different actors in the past is more closely aligned with Koselleck’s project than is sometimes claimed. For Skinner’s point see Quentin Skinner and Javier Fernandez Sebastian, “Intellectual History, Liberty and Republicanism: An Interview with Quentin Skinner”, *Contributions to the History of Concepts*, 3, no. 2 (2007): 103-23: 115. For Koselleck’s account of his own method see Javier Fernández Sebastián and Juan Francisco Fuentes, “Conceptual History, Memory, and Identity: An Interview with Reinhart Koselleck,” *Contributions to the History of Concepts,* 1 (2006): 99–127. Jan-Werner Muller argues something similar, although his emphasis suggests that the history of *concepts* is slightly misleading: noting of Skinner and Koselleck ‘both in a sense agree that concepts do not actually change at all; what changes is the usage of words…’. See Jan-Werner Müller, “On Conceptual History”, in *Rethinking Modern European Intellectual History*, Darrin M McMahon and Samuel Moyn, eds. (Oxford: Oxford University Press, 2014), 74-93, here 87. [↑](#footnote-ref-12)
13. See Hans-Jorg Rheinberger, *Toward a History of Epistemic Things: Synthesizing Proteins in the Test Tube* (Stanford, CA: Stanford University Press, 1997). [↑](#footnote-ref-13)
14. There are some aspects to the field of inquiry, so defined, that remain sketchy or opaque. What theory of language, for example, would support the basic outlines of this method? Strictly speaking words have meanings, or to put that in a more neutral way, words have meaning *effects*. Such an account is based upon a very long tradition of debate in the philosophy of language which tries to keep representation and reference apart as two distinct categories. But does the same model apply to concepts? Although one can make sense of the proposition, say, that buildings have meaning, one can only do so by invoking a set of transfers or translations, that is by utilising metaphor. To say that concepts have meaning is to replicate this move unless one assumes that words and concepts are indeed the same kind of thing, in which case there would be no need for a different theory of meaning and the ontological difference between words and concepts would immediately collapse. [↑](#footnote-ref-14)
15. Here our approach shares some similarity to recent work in cultural theory and history in which function moves to the front of attention. See for example Isabelle Stengers’s notion of ‘operations of propagation’ and ‘operations of passage’ which are said to be properties of ‘nomadic concepts’. These are very close to some of the functions we believe we have identified as outlined in the fifth section of this paper. See Isabelle Stengers, ed, *D’une science a l’autre: Des concepts nomades* (Paris: Editions du Seuil, 1978), 17 ff., 24. [↑](#footnote-ref-15)
16. Jan Ifverson, “About Key Concepts and How to Study Them”, *Contributions to the History of Concepts*, 6, no. 1 (Summer, 2011): 65-88: 69. [↑](#footnote-ref-16)
17. The hypothesis here is that we may be able to discern a ‘grammar’ for conceptual operation that is not entirely dependent upon semantic value and hence ‘parse’ it. [↑](#footnote-ref-17)
18. We are encouraged to note that a similar turn to digital methods is proposed by Phillip Sarasin in this journal, even if his model keeps close to the semantic. See Philipp Sarasin, “Is a “History of Basic Concepts of the Twentieth Century” Possible? A Polemic”, *Contributions to the History of Concepts*, 7, no. 2 (Winter, 2012): 101-110: 108-110. [↑](#footnote-ref-18)
19. In this paper we restrict our few examples to Eighteenth Century Collections Online since the aim is to establish the outlines of a data driven approach as exemplary for further work that might connect to more standard conceptual history. [↑](#footnote-ref-19)
20. Here our approach can be compared to the Linguistic DNA project. See Susan Fitzmaurice et al., “Linguistic DNA: Investigating Conceptual Change in Early Modern English Discourse”, *Studia Neophilologica*, (2017): 1-18. [↑](#footnote-ref-20)
21. See C. Kelly, B. Devereux, and A. Korhonen, “Acquiring human-like feature-based conceptual representations from corpora,” *Proceedings of the NAACL HLT 2010 First Workshop on Computational Neurolinguistics,* Association for Computational Linguistics (2010): 61-69; and F. Pereira, M. Botvinick, and G. Detre, “Using Wikipedia to learn semantic feature representations of concrete concepts in neuroimaging experiments,” *Artificial Intelligence* 194 (2013): 240-252. [↑](#footnote-ref-21)
22. See D.M. Blei, A.Y Ng, and M. I. Jordan, “Latent dirichlet allocation,” *Journal of Machine Learning Research* 3 (Jan. 2003): 993-1022. [↑](#footnote-ref-22)
23. See K.M. Quinn, B.L. Monroe, M. Colaresi, M.H. Crespin, and D. R. Radev, “How to analyze political attention with minimal assumptions and costs,” *American Journal of Political Science* 54, no. 1 (2001): 209-228; and J.M. Hughes, N.J. Foti, D.C. Krakauer, and D.N. Rockmore, “Quantitative patterns of stylistic influence in the evolution of literature,” *Proceedings of the National Academy of Sciences* 109, no. 20 (2012): 7682-7686. [↑](#footnote-ref-23)
24. Weaver first suggested this kind of similarity as a method for word translation in W. Weaver, “Translation,” *Machine Translation of Languages* 14, (1975): 15-23. And for an overview of word-context models see W. Lowe, “Towards a theory of semantic space,” *Proceedings of the Twenty-Third Annual Conference of the Cognitive Science Society*, (Mahwah, N.J.: Lawrence Erlbaum Associates, 2001), 576-581. [↑](#footnote-ref-24)
25. See K. Lund, K. and C. Burgess, “Producing high-dimensional semantic spaces from lexical co-occurrence,” *Behavior Research Methods, Instruments, & Computers* 28, no. 2 (1996): 203-208; D. Lin, “Automatic retrieval and clustering of similar words,” *Proceedings of the 17th international conference on Computational Linguistics* 2, (1998): 768-774; O. Levy and Y. Goldberg, “Dependency-Based Word Embeddings,” *Association for Computational Linguistics* 2, (2014): 302-308. [↑](#footnote-ref-25)
26. See T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient estimation of word representations in vector space,” *arXiv*, (2013): preprint arXiv:1301.3781; D.R.G.H.R. Williams and G.E. Hinto, “Learning representations by back-propagating errors,” *Nature,* 323 (1986): 533-536. [↑](#footnote-ref-26)
27. For example, Magnus Sahlgren, *The Word-Space Model: Using distributional analysis to represent syntagmatic and paradigmatic relations between words in high-dimensional vector spaces*. PhD dissertation, Department of Linguistics, Stockholm University 2006. [↑](#footnote-ref-27)
28. See M.A. Hearst, “Automatic acquisition of hyponyms from large text corpora,” *Proceedings of the 14th Conference on Computational Linguistics*, Association for Computational Linguistics, Volume 2, (1992): 539-545. [↑](#footnote-ref-28)
29. See R. Girju, A. Badulescu, and D. Moldovan, “Automatic discovery of part-whole relations,” *Computational Linguistics* 32, no. 1 (2006): 83-135. [↑](#footnote-ref-29)
30. Note that we do not lemmatize the corpus in order to retain important distinctions in abstraction often indicated by different word forms, for example *despot* vs *despotism*, or *nerve* vs *nerves*. [↑](#footnote-ref-30)
31. We note that approaches which embed co-occurrence based measures to discriminate between paradigmatic and syntagmatic relations are possible and could be a worthy future direction to pursue.

 [↑](#footnote-ref-31)
32. See Lukas Michelbacher, Stefan Evert, and Hinrich Schütze, "Asymmetry in corpus-derived and human word associations", *Corpus Linguistics and Linguistic Theory, 7,* 2 (2011), 245–276. [↑](#footnote-ref-32)
33. See M. Setnes, R. Babuska, and H.B. Verbruggen, “Rule-based modeling: Precision and transparency,” *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 28, no. 1 (1998): 165-169; J.D. Olden and D.A. Jackson, “Illuminating the “black box”: a randomization approach for understanding variable contributions in artificial neural networks,” *Ecological Modelling* 154, no. 1 (2002): 135-150; H. Luo, Z. Liu, H. Luan, and M. Sun “Online learning of interpretable word embedding,” *Proceedings of EMNLP* (2015): 1687-1692. [↑](#footnote-ref-33)
34. See K.W. Church and P. Hanks, “Word association norms, mutual information, and lexicography,” *Computational Linguistics* 16, no. 1 (1990): 22-29; J. Pennington, J.R. Socher, and C.D. Manning, “Glove: Global Vectors for Word Representation,” *EMNLP* 14 (2014): 1532-43. [↑](#footnote-ref-34)
35. See C. Manning and H. Schütze, *Foundations of Statistical Natural Language Processing* (Cambridge, USA: MIT Press, 1999). [↑](#footnote-ref-35)
36. For examples, see Pantel, Patrick, and Dekang Lin. "Discovering word senses from text." In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, 2002: 613-619. [↑](#footnote-ref-36)
37. See C. Manning and H. Schütze, *Foundations of Statistical Natural Language Processing* (Cambridge, USA: MIT Press, 1999). [↑](#footnote-ref-37)
38. For how the measure described in Section 3.2compares with similar measures in this regard, see Gabriel Recchia and Paul Nulty. "Improving a fundamental measure of lexical association." In *Proceedings of the 39th Annual Conference of the Cognitive Science Society (Austin, TX, 2017)*. Cf measure in 3.2 above described in Recchia & Nulty as “Simple PMI”. [↑](#footnote-ref-38)
39. O. Levy, Y. Goldberg, and I. Dagan, “Improving distributional similarity with lessons learned from word embeddings,” *Transactions of the Association for Computational Linguistics* 3, (2015): 211-225. [↑](#footnote-ref-39)
40. See A. Søgaard, A. Johannsen, B. Plank, D. Hovy, and H.M. Alonso, “What’s in a p-value in NLP?,” *Proc. CoNLL,* (June, 2014): 1-10. [↑](#footnote-ref-40)
41. See D.A. Jackson, “Stopping rules in principal components analysis: a comparison of heuristical and statistical approaches,” *Ecology* 74, no. 8 (1993): 2204-2214. [↑](#footnote-ref-41)
42. In order to preserve the data on raw frequency we report this data in brackets following the specified word. [↑](#footnote-ref-42)
43. Edmund Burke, *Reflections on the Revolution in France*, ed. Conor Cruise O’Brien (London: Penguin, 1968), 86. [↑](#footnote-ref-43)
44. *The Works of Edmund Burke with a memoir*, 3 vols.(New York: George Dearborn, 1835), I, 201. [↑](#footnote-ref-44)
45. See Michael Hoey, *Lexical Priming: a new theory of words and language* (London: Taylor and Francis, 2005). [↑](#footnote-ref-45)
46. See Thomas M.J. Fruchterman and Edward M. Reingold, ‘Graph drawing by force‐directed placement’, *Software: Practice and Experience,* 21.11 (1991): 1129-1164. [↑](#footnote-ref-46)