DHQ: Digital Humanities Quarterly
2022
Volume 16 Number 4

Interpreting Measures of Meaning: Introducing Salience Differentiated Stability

Hugo Dirk Hogenbirk <h_dot_d_dot_hogenbirk_at_rug_dot_nl>, University of Groningen, Departement of the History of Philosophy
Wim Mol

Abstract

In digital studies of the use of words in intellectual history, meaning is measured based on the idea of Firth that a word can be characterized by the company it keeps. The words that are literally close to it in the texts in which it is written should tell us something about what the word means. In practice, we will look at meaning being measured by a method called *Pointwise Mutual Information* or PMI for short. However, even granting that we use PMI, this description is still quite vague and allows for multiple different ways to interpret 'the company a word keeps' in the practice of coding an actual algorithm to discern this company. In this paper we will look specifically at the choice to 1. use all words close to another or 2. use merely the ones that are most disproportionately present. Using work from contemporary philosophers of language Mark Wilson and Sally Haslanger, we argue that both capture an aspect of meaning, 2 capturing the most salient way to understand a word, and 1 capturing the subtle, not so salient, but nonetheless important ways in which words are used overall. Then we will look to measure the overall stability in word meaning, the degree to which it is used similarly in different texts within a corpus. Having characterized PMI, salience and stability we will introduce *Salience Differentiated Stability* or SDS, a value indicating both the salient and less salient stability of a word, which will help identify words that are simpler or more shifty than they at first appear. Lastly we will test the use of this new value by doing a case study of the salience differentiated stability of common terms in early modern physics text books.

Section 1: Introduction

In digital humanities it is common to study the change in use of a word in a corpus of text by means of vector semantics, or distributional semantics. The idea is to map word-types found in a text or in an entire corpus to vectors.[1] Each coordinate of the vector will represent an association with another word-type. An algorithm can be written to scour parts of text near the word-type for associations. The idea is that a word like, for example, 'food' will automatically be associated with other words that tell us about the concept of food. Prominent examples of food, like 'cheese' or 'bread', or near synonyms like 'nutrition', or biologically related organs like 'stomach'. Such methods have at least the following advantages: first, they are fully automated, and hence, large corpora of text that could not realistically be studied manually can now be studied. Secondly, they can track homonymy, (near) synonymy and changes in the use of an investigated word, even as the word stays linguistically constant. Linguistically these methods are justified by the idea due to John Firth that “You shall know a word by the company it keeps” [Firth 1957]. Firth, and somewhat synchronously, Zellig Harris, defended and developed the view that the verbal connotations of a word gave us important insight into its use, more so than any definition could do [Firth 1957] [Harris 1964]. We need not go that far. To justify the idea of vector semantics we only need to believe that the distributional properties of a word give us some semantic insights into the word that are of interest to a purveyor of large corpora [Gavin 2019]. 1

However, this does not settle the exact details of our vector semantics. There are many different ways to calculate the exact values of each of the coordinates, and it is not clear which of these characterizations of the company of a word tells us what. Some people might assume that one or the other characterization gets us closer to a word’s meaning, to be expressed for example, by scoring higher or lower on pre-defined metrics. We will offer an alternative to this approach by arguing that different methods might give us different facets of meaning. 2

In this paper we look at two particular ways to turn the company of a word into vectors. One, where each word-type in a corpus is mapped to a number representing the frequency with which it occurs near another word-type relative to its overall frequency, and another where a word-type is associated with a list of high scoring words. Both of these methods find the 'company' of a word in some way. However, we will claim, these results need not be the same for a given word-type. Hence, the main purpose of this paper is to offer an interpretation of what these different ways of specifying ‘the company' tell us, so that an investigator can i) more easily choose between the two methods and ii) use them both next to each other profitably. Both of these goals constitute a novelty; i) depends on the interpretations of the methods as providing different (and not competing) facets of meaning (sec.2). By doing such analysis we are applying insight from recent innovations in the philosophy of language that have come to fall under the name of ‘conceptual engineering’, to the interpretation of language by algorithms. ii) is cashed out in a new method that measures the divergence of the two methods as a new semantic measure (sec.4) whose object has no counterpart in the literature. In order to interpret these methods we will examine their workings in detail. 3

In section 2, we draw on the work of philosophers like Sally Haslanger and Mark Wilson to argue that a word has salient, more publicly known and obvious semantic content as well as more subtle and hence less salient content. We will argue that both are important aspects of the semantics of a word. The subtle, less salient, content is usually understood as an ability to navigate subtleties which we cannot or would not make explicit. We will argue that our restricted-list method tracks the salient connotations whereas the unrestricted method tracks its more complete use, including hidden semantic subtleties. 4

In section 3 we need to dive into the technicalities of the methods discussed. We will show in detail the operational difference between measuring more and less salient semantic content. In addition we will introduce methods to measure the overall stability of word-meaning in a given corpus. Stability will 5 be a measure of the similarity of word use across multiple texts. Lastly we will introduce our novel methodological proposal which we dub Salience Differentiated Stability, which is a two-vector signifying a word's relative salient and less salient stability. This is a metric that will be applicable to any given corpus of texts.

In section 4 we provide an interpretation of SDS-scores. A word can score high and low on both salient and non-salient stability. We give interpretations for all four possible cases. 6

Finally, in section 5, we provide the results of an explorative case study where the Salience Differentiated Stability of a small set of word-types in a corpus of 17th and 18th century physics/natural philosophy is measured and analyzed. This is not a history paper per se, but rather an examination of methodology. Hence, the purpose of this case study is not to study 17th century physics/natural philosophy. It is rather to see whether the methodology, which this paper is about, yields results that make prima facie sense and are non-trivial. They might be trivial if varying salience yielded no difference in results or none that could be plausibly interpreted. We will find in the case study that words which score high on salient stability, do not necessarily score high on subtle/less salient stability, or vice versa. This means that in actual historical examples, salient and less salient semantic content can be quite different both in content and in overall stability. 7

Section 2: Salience

When it comes to vector semantics, there are different ways to code the algorithms that determine the semantic content of a word in a given body of text. We will look at two methods using PMI and vector embedding, where a word is characterized as a large vector, each of the coordinates in the vector signifying how often another word-type appears in close proximity to it. The number appearing in these coordinates is called the collocation score. So, if we are characterizing 'gravity' and 'gravity' and 'force' often appear close to one another in a text, the coordinate signifying 'force' will have a large number (see section 3 for the relevant math and coding). The main methodological choice we will be looking at is the following: when constructing a vector for a word we might decide to ignore all but the most extreme coordinates, that is, all but those words with the highest collocation scores, but we don't have to restrict our attention in this way. We could instead look at all of the coordinates in our vector. The question is: which of these ways of coding best captures the semantic content we want to capture. One way to answer this is by arguing that the meaning of a word is mostly determined by the words it is associated with most or alternatively, the way it is associated with all other word-types in the text. We believe both these answers can coexist in a way. In this section we will argue that there are different ways to conceive of the semantic content of a word that justify the use of different algorithms. 8

The idea of distinguishing different kinds of semantic analysis already exists in the work of feminist philosopher Sally Haslanger. In her *What Are We Talking About? The Semantics and Politics of Social Kinds* she distinguishes different forms of analysis that yield different facets of the meaning of the same word [Haslanger 2012, 365–380]). A conceptual analysis yields the manifest meaning, the way we (or for our purposes, perhaps not us, but some historical population) explicitly understand the meaning of a word. A descriptive analysis would yield the use of the word which might in fact diverge from the way we think we use it. Haslanger gives the example of her son’s primary school which makes use of a concept of tardiness. Officially, any child that arrives after 8:25 was tardy but as Haslanger's son pointed out: “Don’t worry Mom, no one is ever tardy on Wednesdays because my teacher doesn’t turn in the attendance sheet on Wednesday until after the first period” [Haslanger 2012, 268]. Here the use of 'tardy' diverges from the official, explicit definition. Its manifest meaning, which you would have learned by asking the teachers or school staff what ‘tardy’ is, was not the operative meaning, which you learned from studying the tracking mechanism. Beyond this rather local example, Haslanger is thinking about concepts of race and gender, 'man', 'woman', 'white', 'black', etc. These concepts might superficially, or manifestly, be thought to refer to biological categories. For this reason we might argue that the whole concept of race is misguided since the underlying biological category either does not exist or is not nearly as pronounced as the word implies. However, there is also a de facto use of our race categories that deserves study – because even though the manifest meaning might turn out to be bogus, this does not mean these sorts of word are not still doing relevant and investigable semantic work, to be found by looking at the operative meaning [Haslanger 2012, 221–247]. 9

In his book *Wandering Significance* [Wilson 2006] Mark Wilson uses the concept of a façade to denote concepts that superficially seem univocal but on closer inspection have subtly divergent uses in different contexts [Wilson 2006, 147]. A simple example is that of a rainbow [Wilson 2006, 21]. In some contexts it makes sense to say a pot of gold is at the end of it, presumably to be reached by riding a unicorn over it. In other, more literal contexts, rainbows have no clear ends, and cannot be ridden. 10

A better, but more complicated example is that of hardness. Most of us have some notion what it is for a material to be hard, but upon inspection of the technical details, the concept reveals itself as something much more heterogeneous than it at first appears. Wilson notes that a range of different tests and metrics exist depending on what type of material specialists speak of and for what purpose. Some of these tests identify hardness with resistance to scratching or cutting and some of these metrics identify it with flow stress or a resistance to penetration. Wilson gives examples: “In these contexts [indenter tests for common metals], yield strength becomes the attribute to which “hardness” locally gravitates (although such identification is completely unnatural for glass)” [Wilson 2006, 341]. As Wilson notes, not all of these tests even make sense for each type of material and for each purpose. Hence hardness relates to a range of local patches of meaning that are tied together by the vague general notion of hardness but are each individually quite different. The use of the concept admits of a lot of subtleties that are not apparent at face value, often because they only appear in exotic contexts or because different specialized contexts are usually kept separate. This means that we should expect the use of the concept to be more divergent than its most salient description. Another example is given by Thomas Kuhn [Kuhn 2012, 188] who argues that the well-known formula *F=ma* is applied quite differently in different branches of physics. 11

We agree with these authors on the following points: first that the use of a word is not always the same as the explicitly given definition or the first description of it that comes to mind. Secondly that the use of a word often has a lot more subtlety and divergence to it than we commonly give credit for. Thirdly that both the subtle uses of the word and its more salient, obvious meaning are of interest. We will use the notion of salience to distinguish greater and smaller degrees to which the meaning of a term will be public and obvious to its users (authors and readers). The idea is that the more 12 extreme coordinates of our vectors will show us the more salient aspects of a word's meaning, whereas the smaller coordinates and the differences between them are still relevant to the way a word is broadly used, but not in a way that readily comes to mind even to those who use the word regularly. Unlike different types of meaning or analysis, salience will not be a binary division with salient and non-salient features but will admit of degrees. This suits our purposes because in the digital methods employed we will distinguish the salient from the not so salient by a continuous variable, occasionally using an unavoidably arbitrary cut-off point of a value above which word-relations will be considered salient. Hence, 'salient' will be used much the same as 'large', supervening on an underlying property that varies continuously (size) and being only understandable in relative terms (large compared to a human is different from large compared to the sun).

Section 3: Algorithm implementation and details

Our two methods for the extraction of semantic information about word-types from a corpus are the following: (1) extraction of lists of collocates and (2) generation of a vector-representation. We will first introduce the generation of a vector-representation of a word in a given text and then argue that extracting collocates is a specific restriction on this vector-representation that allows it to bring to the fore the more salient features of a word-type. Then the different ways of measuring similarity between word-types will be introduced and the particular application of our proposed algorithms for conducting our case-study. Two words are of the same word-type if they are spelled identically. We might consider grammatical cases of the same stem as identical as well, in which case the unit under discussion is a lemma rather than a word-type. Word-types are used throughout as the objects of study, however, for the case-study in Section. 5 we've lemmatized the texts and are thus looking at lemmas. However, the methods proposed are equally applicable to both types of object. 13

Vector semantics

In accordance with Firth's and Harris' intuitions, the algorithm works by considering the contexts of word-types (by looking at what word-types surround them) and saving that data in a proper data-structure – meaning that word-types will be represented *as* the other words that they co-occur with and the counts of these co-occurrences. First we must get the base information we need from the texts. This is done by counting for every word-type which other words co-occur with them within some window of size n. Here n signifies the number of words that we look at on both sides of all the occurrences of the word-type investigated. This means that if we for example choose a window of size 4, and consider the following sentence: 14

> The movement towards digital hermeneutics is fraught with difficulties, but movement is never without difficulties.

We can derive a representation of this sentence in terms of relative closeness of word-types. There are 12 different word-types in the text. For each word-type we can ask how often each word-type occurs (given a particular windowsize). In table 1 below you can see what values this would deliver (the first column shows how many occurrences of the word-type (WT) occur in total): 15

#	WT	The	movement	towards	Digital	hermeneutics	is	fraught	with	difficulties	but	never	without
1	The	1	1	1	1	1	0	0	0	0	0	0	0
2	Movement	1	2	1	1	1	2	1	1	2	1	1	1
1	Towards	1	1	1	1	1	1	1	0	0	0	0	0
1	Digital	1	1	1	1	1	1	1	1	0	0	0	0
1	Hermeneutics	1	1	1	1	1	1	1	1	1	0	0	0
2	Is	0	2	1	1	1	1	1	2	3	2	1	1
1	Fraught	0	1	1	1	1	1	1	1	1	1	0	0
1	With	0	1	0	1	1	2	1	1	1	1	0	0
2	Difficulties	0	2	0	0	1	3	1	1	1	1	2	1
1	But	0	1	0	0	0	2	1	1	1	1	1	1
1	Never	0	1	0	0	0	1	0	0	2	1	1	1
1	Without	0	1	0	0	0	1	0	0	1	1	1	1

Table 1. Example of construction of a table of all co-occurrence vectors

This is a vector-representation of the contextual information in the text. Each of the rows is a vector (i.e. there is a vector for each word-type) in a 12-dimensional space. Every individual number in the table represents how often the column word co-occurs with the row word. Important to note is that the windowsize directly impacts the nature of the results of the vector-representation of the text – in our experiment in section 5, we use a windowsize of 10. There is a strong case to be made that taking different values for the windowsize can in many ways be used to extract not better or lesser results, but properly different results from the text.[2] 16

However, these raw counts are, on their own, not particularly informative. They need to be transformed so that we can extract values that indicate how strongly two word-types are connected that is not directly influenced by the total occurrences of certain word-types. We see in the above table for example that 'movement', 'is', and 'difficulties' all have higher total co-occurrence counts than the other words – a direct consequence of their own higher frequency. However, it is not so that by occurring more often, a word-type is necessarily more connected to more types of words in interesting or salient ways. This means we need to transform these raw counts-scores into a score that is unaffected by total frequency of word-types. For this we have used and will now introduce the measure *pointwise mutual information.* 17

Measuring: PMI

18 Each of the pairs of word-types needs to be scored – or, their connectedness needs to be measured – based on the data from Table 1. The way to do this is by extracting the data about the number of times they have co-occurred, together with how many instances of the word-type there were. The method we will be using to do this is *Pointwise Mutual Information* (PMI).[3] PMI measures the chance of finding a word-type *y* within a window around another word-type *x* (within a chosen windowsize) in the investigated text, divided by the chance of finding y in the whole text. The idea is that related words should be found disproportionately in each other's windows[4]. The overall chance of finding *y* indicates a regular proportion. We take the log of this value in order to make the results better suited for vector calculus later. We will write small *x* for the set of instances of x itself and capital *X* for the windows around instances of *x*. Notice that the different windows *X* might overlap. In this case the same word-token will be counted twice or more. Hence *X* should be seen as the disjoint or indexed union of the windows. The formula for PMI reads:

19 $$PMI(x,y) = log_2 \frac{P(y \in X)}{P(y)}$$

20 This provides us with the average of the chance of finding a token of *y* in a window near a token of *x*, divided by the overall chance of finding *y* in the text. The fraction can only be a positive value (both denominator and numerator are positive), but a log of a value between 0 and 1 gives a negative value. If the two word-types are fully independent from each other, we'd see that there is no difference in the chance of finding *x* and *y* in window-proximity to each other in our actual text, versus the hypothetical situation where they have no actual relation – i.e. the fraction will then give as a result 1. *Log(1) = 0*, so PMI scores above 0 should be read as indicating a positive correlation between the two word-types in the text, whereas scores below 0 indicate negative correlation, and 0 itself indicates independence.

21 Computationally the chance *P(y)* is simply the instances of *y* divided by the total word count. The chance P$(y \in X)$is found by counting each instance of *y* within the window of an instance of *x*, where the same *y* may be counted more than once when it appears in more than one window, divided by the size of the disjoint union *X* of all the windows, which will always be the window size multiplied by the amount of times *x* appears in the text.

22 $$\frac{p(y \in X)}{P(y)} = \frac{\left(\frac{|y foun_d inwin_d owsaroun_d x|}{|x| * windowsize}\right)}{\left(\frac{|y|}{Tota_l wor_d s}\right)}$$

23 So for example, when we consider table 1 above, we can see that the score for PMI(is, with) with a window size of four is given by the fact that we find 'with' around 'is' twice, that 'is' occurs twice itself, that 'with' occurs once and that the entire text consists of 15 words, giving us:

24 $$log2 \frac{\frac{2}{2*4}}{\frac{1}{15}} = \left(\frac{0.25}{0.0667}\right) = 3.75$$

25 When the corpus is only a single sentence, this method does not yet pick up on any significant semantic properties – but as we will see later on, when the number of sentences and words, grow, this problem diminishes.

Divergence: extracting collocations and the salient

26 From this vector-representation of a word with scores based on PMI, there are (at least) two routes available for the extraction of different meaning-representations. We argue that one discovers more salient aspects of meaning, whereas the other discovers the more subtle/less salient aspects of meaning.

27 Many of the scores inside the vector representation are low. The connectedness of the terms is there, but not particularly salient. Specifically within the approach that is called *the extraction of collocations*, this intuition is put to good use. Collocates to a specific word are those words that are *particularly connected* to the word that is being investigated. Instead of looking at 100.000 word-types and their connectedness to a word of interest, in these cases, one restricts the list to those word-types that score at least a certain score on the PMI measure. This often leaves investigators with a handful, up to a ~100 word-types, which together provide a good overview of the interesting semantic properties of the word-type under investigation.[5] In our own algorithm, we have taken the threshold to be a PMI-score of 4, so as to mimic these applications, which often aim to consider around a few dozen of word-types for an investigated word-type.[6]

28 We argue that what investigators making use of lists of collocates are extracting are those connections between word-types that are (within the investigated corpus) particularly salient, and by extension, more public and explicit. This works well for their purposes. The list of high scoring words will mostly include words that, to a reader of, or author in, the corpus, will likely 'make immediate sense' when provided. Of course, the nature of the connection won't be provided, but that they are connected will be clear on most occasions to most people who read the corpus – and looking at this limited set of words can provide an investigator with a lot of useful information on the corpus.

29 Contrast this with the vector representation. Here the meaning of a word-type is not given by a list of *x* particularly saliently connected words, but by the PMI score with every single other word-type contained in the corpus. In addition, instead of taking an implicit 'one-hot' representation as in the collocation extraction (another word-type either is, or is not, connected enough) the vector representation typically takes the exact outcome of the PMI calculation as its score. This means that even minute differences will play a role in the differentiation between different word-types. What we are picking up on here, we argue, are the more hidden, less explicit, connections a word-type makes within the corpus. It investigates how the particular ways of using a term influence further usage and application of a term, despite it being particularly hard for a speaker or reader to explicate that such a causally productive connection exists.

30 We have suggested that PMI scores track relevant semantic properties of a word. In addition, that high scores track more salient features, and low scores more implicit and subtle features. What argument or evidence is there for this? Perhaps the use of a word is part of its meaning but surely what other words it appears near in written text is at best an approximation of how the word is used.

31 Psychological research shows that PMI is positively correlated with judgment of similarity, that is to say if words have higher PMI scores in sample texts then test-subjects are more likely to judge them similar [McDonald 2000, 35–67]. In addition, texts with high PMI scores for given word pairs have been shown to make people see the words as more related in meaning [McDonald and Ramscar 2001]. In addition, as McDonald points out, such scores can be correlated with other psychological variables such as effect in lexical priming [Lowe and McDonald 2000], analogous reasoning [Ramscar and Yarlett 2000] and synonym selection [Landauer and Dumais 1997].

32 This shows that whatever properties are tracked by PMI are not mere artifacts of the texts studied but have at least some correlation with psychological indicators of word meaning. What does not follow from this work is that lower PMI scores indicate less salient aspects of the use of a word. That would, by its very nature, be hard to extract from straightforward questions to test subjects, though it might be studied in other ways. In section 5 we will include results from an explorative case study that show that in at least some cases, looking at just the high scoring words yields different results from looking at all word associations.

Similarity measures

33 The above methods describe how we propose to be able to extract an approximation of less or more salient aspects of meaning. In the first case, the vector is the representation of the less salient aspects of the concept of a word-type in a text/corpus, in the latter case, the list of connected terms is the representation of the more salient aspects of the concept of a word-type in a text/corpus. In the case of the extracted list of collocates, research sometimes focusses on the qualitative investigation of this list of words itself (indeed, the ability to do so is one of the attractive features of the entire approach).[7] The vector approach generally does not allow for this and is often coupled together with a measure of the similarity between vectors. This allows the researcher, for example, to check the stability of terms over time, find synonyms, and discover dissimilarity between terms which are expected to hang together.[8] We introduce a commonly applied similarity measure for vector-representations and argue that there are useful options for measuring the similarity of lists of collocates. These measures will be used in the case study to extract useful information about our corpus for both sorts of meaning-representations.

34 In the case of the vector representation, what we are looking for is actually a measure for *closeness* of coordinates in a multi-dimensional space. A number of different options is available, but one of the more often used measures is *cosine similarity* [Han 2012, sec. 2.4.7]. Intuitively, how similar two vectors are is measured by looking at how much they coincide for each of their coordinates. A simple measure for this is provided in linear algebra by the inner product, or, dot-product. This measure sums the products of each pair of coordinates that lie in the same dimension, or:

$$\vec{v} \bullet \vec{w} = \sum_{i=1}^{N} v_1 {}^* w_1 + v_2 {}^* w_2 + .. + v_N {}^* w_N \quad 35$$

36 Partly, this measure does what we want; the more (high) values shared between two vectors in similar dimensions, the higher the similarity of the two vectors will be. However, what is counter-intuitive is that longer vectors (vectors with higher values in many dimensions) will *generally* be scored higher than shorter vectors will be. However, it is not the case that just because a word-type has more strong connections with other word-types (higher PMI-scores) it should be deemed to be expected to be more similar to other word-types in general. This problem can be amended simply by normalizing the lengths of the two vectors; in this way, we only measure the similarity in distribution of values across the different dimensions of the two vectors, while negating the influence of their absolute lengths. We get for our similarity measure:

$$\frac{\vec{v} \bullet \vec{w}}{\vec{v} \; {}^* \; \vec{w}} \quad 37$$

38 Linear algebra shows that this formula captures the same thing as taking the cosine of the angle between the two vectors *v* and *w*, hence the measure's name *cosine similarity*. Higher values are attained when the distribution of values between two vectors amongst all the possible dimensions is similar – or, when the angle between the two vectors is close to zero (in addition, for orthogonal vectors, we get a similarity of 0, and for opposite vectors, a value of -1). This allows our measure to provide values between -1 and 1.

39 In the case of the lists of terms, how to model similarity is less researched because lists of terms do not easily translate into a question about vectors in space. Intuitively, however, what we want to measure is the amount of overlap two terms exhibit in their lists of collocates. The greater the overlap, the larger the amount of saliently connected terms two words share. What is dangerous however is that we do not want that word-types that have a larger number of strongly connected terms be more similar in general to other words (which would happen if we would take 'a large overlap' to be signified by a large number of overlapping terms). This needs to be normalized. One of the measures used for these purposes is the Jaccard index, which scores the overlap between the neighborhood of two nodes in a network in the following way:

$$\frac{|A \cap B|}{|A \cup B|} \quad 40$$

41 The intersection provides the overlap and the union provides the sense of the 'total' size of the neighborhoods; sharing a large amount of terms provides a higher score the smaller the total sample size of potentially shared terms becomes.

42 This measure is straightforward in the special case that the neighborhoods have the same cardinality. The sets {E, F, G, H} and {G, H, I, J} will have a Jaccard index of 1/3; the intersection contains two elements (G and H), whereas the union contains six elements (E, F, G, H, I, J). In the case of the same cardinality, we also have the nice property that complete overlap will provide a score of 1. This breaks down in the case of dissimilar cardinalities. Take the Jaccard Index over {E, F} and {E, F, G, H}. This will turn out to be a half (two shared elements and four total elements). This means that the scores will be influenced by the lengths of the lists of collocates. However, this on its own is not a bad thing, as the length of the list tells us something about how much salient connections there are for the word within the corpus. If two terms significantly differ in this respect, it is reasonable to take this along in the calculations. What we do want to avoid (and what the Jaccard Index avoids) is that words with larger or smaller lists of terms get scored more or less highly on their similarity scores in general. This is not the case, since the length of the list only comes into play relatively to the length of

the list of another word-type with which the similarity is measured. Word-types with the same, or a similar, neighborhood size will have a higher potential score, but this does not translate into a preference for either larger or smaller neighborhoods to influence the scoring on its own. In addition, the Jaccard index is commutative. It will always tell us that a first word is as similar to a second word as the second to the first. This maintains the intuitive symmetry of the similarity relation.

Intermezzo: a comparison with word embeddings

As a brief intermezzo we will discuss why our mode of interpretation, which focuses on the salience of individual scores in the PMI-based semantic vector, would be much harder for word embedding based methods like word2vec [Mikolov et al. 2013] and BERT [Devilin 2018]. Such methods have been used in many semantic and linguistic contexts, including recently the tracing of semantic change in historical corpora [Hamilton 2016], [Wevers and Koolen 2020]. In fact, for many tasks like tracking synonymy and homonymy word-embeddings are superior to the sparse vector approaches that have been interpreted above. 43

Word embeddings can make use of different architectures, but let us look at the continuous bag-of-words model. Making use of a neural net, an algorithm attempts to invent word vectors that are maximally useful in the task of predicting, on the basis of a collection of words surrounding a word's position (the bag-of-words), what word will occur in the middle of these words. By altering the values (weights) of a word's entries into its vector so as to be maximally effective at this task, a vector representation of the word is learned. As the algorithm learns, it doesn't find the eventual weights, or scores, of words along a number of axes, it defines its own, maximally informative, axes. That is to say, in using a method like word2vec, not only are the scores that should define a word-vector extracted, but also the best conceptual frame of representing the word-vectors is constructed. However, due to the difficulty in interpreting what a particular axis stands for, it also becomes difficult to interpret individual scores along this axis as is done above (entries being interpreted for example, as more or less salient connection between word-types). 44

Consider the vector describing word type 'The' in table 1. This vector has 12 values, each representing the number of times the word 'the' co-occurred with the 12 other word types. Each of the vectors can be retraced to interpretable quantities through its elements which are interpretable quantities. By contrast a comparable word2vec vector would contain entries that represent variables invented by the algorithm. This does not allow for the word-type-pair-based analysis we propose. 45

To put it a bit more technically, word embeddings use various techniques to represent semantic properties in lowdimensional spaces. This reduction in the amount of dimensions has computational advantages, but the downside from an interpretative point of view is that it is less clear what the elements of vectors represent (in any case, there is no reasons to expect them to represent word-type-pair relations). Our methods also allow us to define a vector-space and norms on it (see the section on similarity measures) which has far more dimensions, but the directions, vectors, and individual entries can be much more readily interpreted because this space was invented by and for people. 46

It might be possible, instead of trying to interpret the individual steps of the algorithm produced by a teaching algorithm, to instead interpret what word embedding methods are good at, or what using certain training data teaches them[9], but that is beyond the scope of this paper. 47

Stability: Introducing salience differentiated stability scores

Having discussed two ways to model words, vector-representations and collocate-representation, which respectively model the less and the more salient aspects of its meaning, we now move to the concept of the stability of a word-type within a corpus of individual texts, which is the final ingredient required for our proposed application of combining differently measuring the more and less salient aspects of meaning. 48

For each of the works in a given corpus we can generate both vector and collocation representations of a set of different word-types that are relevant to the corpus. Using the similarity measures we have discussed above, we are able to compare each of these models to one another. In particular, we compare the models of the same word in different works. For instance, the usage of the term 'cause' in one work of the corpus can then be compared on the level of similarity (for both facets) with the usage of 'cause' in another work. For our two methods we then get a relation of similarity for each word-type between every work in the corpus. These results can be used for building up a network representation of the corpus. Consider for example Figure 1 below, which provides the resulting graph, based on the network of all similarity scores between all works in the corpus of early-modern English physics/natural philosophy textbooks that will be further detailed in section 5, for the vector-representation based method, indexed on the word 'body': 49

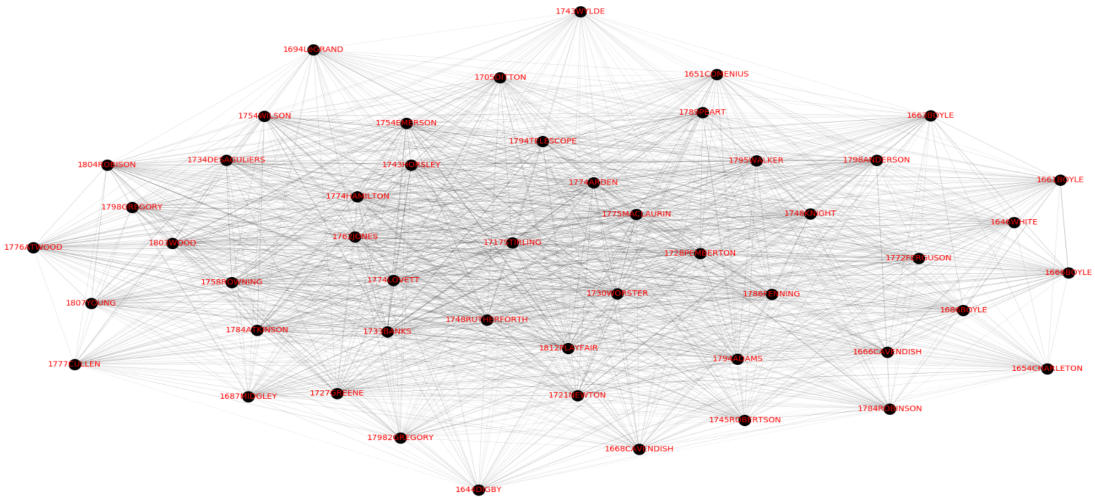

Figure 1. Resulting network representation from comparing each work in the corpus to every other using similarity scores based on their models of ‘body’ that incorporate the less salient aspects of its meaning.

The nodes stand for the different works in the corpus, while the weighted edges that connect them stand for the similarity between the two works in the generated models of the investigated word-type. Of course, one can analyze such a (set of) networks (one for each word-type) in many different ways to allow for interesting corpus analysis.[10] We will only make use of a particular route of investigation – we will look at the average weight of the edges for each of these different networks (i.e. word-types) for both of the models available. On its own, this average of the edge weight intuitively gives a number that indicates how similar the word is used across the corpus. A high score indicates that many of the works have particularly similar models for their particular meaning of the investigated word-type. A low score by contrast tells us that the word is not used particularly similar across the corpus, and that the meaning of the term is particularly unstable. We’ll label these scores ‘*stability scores*’. Higher values signal that a term under investigation is used more similarly across the entire corpus, i.e., the term’s usage is more stable in that corpus than one that scores lower. 50

Our important claim regarding these numbers comes from the divergence between stability scores obtained by the same word-type, depending on whether we are considering their more or less salient aspects of meaning. Indeed, whereas one might expect to find two kinds of terms – terms that have high stability scores for both the more and less salient aspects of their meaning, and that score low on both – we find divergence between these scores. It is these divergencies of stability scores when one differentiates between more and less salient aspects of meaning that must be interpreted and that will be provided as our case-study. To investigate such scores we call *salience differentiated stability analysis*. 51

Section 4: A framework for interpreting SDS-scores

We have used two methods to distinguish similarity between word usage. One of them, we argued, measured the most salient associations with a word, whereas the other measured it’s overall usage, including the most subtle differences in emphasis. We then used the Jaccard index and cosine similarity to gain an overall measure of how much a word is used similarly across texts, according to our two methods. It turns out that the two scores are not always correlated. Some words score highly on their average jaccard score but low on their average cosine similarity and vice versa. If this were to happen often, and in ways we could predict this would in itself be evidence that they measure different things. 52

For doing salience differentiated stability analysis, we can divide words into four categories, or even better, a two-dimensional grid where each word is closer to one corner or another. A word will either have a lot of salient difference in usage or not across a corpus, and it will have a lot of less salient differences in usage or not. Something interesting is to be said of each of these four categories. 53

- The simplest are words which score high on both scores. These are words which are similarly understood by most authors and which are used in much the same way. We should expect them to have a straightforward and broadly understood meaning. Given that the terms are central to the corpus (to be ascertained via domain knowledge, or other methods, like topic modelling) one can expect the terms to be discussed, but not conceptually controversial.
- Words which score lowly on both scores are also quite straightforward. These are words which are understood differently and used differently. We should expect them to be at the center of controversies which are difficult to resolve, or to be outright polysemic.
- More interesting are words which are similar in salient ways but dissimilar in more subtle ways. Here superficial similarity may mask slippery differences in usage. We will call such words façades. If we follow Haslanger we should expect these words to be used in masking certain hypocrisies and if we follow Wilson these words may suggest a straightforward usage when they are in fact used in a façade-like manner, being used subtly differently in different areas of application. Alternatively it could indicate a controversy in a narrow domain, where the broad subject of discussion is fixed but the details are the subject of controversy.
- Lastly we have words which have a lot of salient dissimilarity but which are overall used relatively similarly. We should expect these words to be the subject of controversy whilst their actual use is relatively fixed.

Section 5: Explorative case-study of early modern natural philosophical terminology

Now we turn to the case study. We reiterate: the purpose of this case study is not primarily to study the history of natural philosophy/physics. It is rather to provide a proof of concept of salience differentiated stability. If the case study were to bear out that all words are similarly stable, or that Jaccard scores do not yield different results from cosine similarities, our methods would be trivial. If some of the results were not plausibly interpretable, the results would simply be too garbled to show anything meaningful. Neither appears to be the case in this case study which provides at least a minimal proof of concept for salience differentiated stability. 54

Our corpus consists of early-modern British natural philosophical/physics textbooks (1600~1800), published in English. The corpus has been built within the context of the “The Normalisation of natural philosophy” project, based at the university of Groningen. It consists of abroad selection of natural philosophical/physics texts, annotated for their level of systematicity, from four geographical areas (Great-Britain, The Netherlands, Germany and France), between 1600 and 1800.[11] From this corpus we’ve made use of the most systematic works in the British English language corpus that were also available in a digitized format. This led to a sub-corpus of 50 works, each of which has been OCR’ed in the context of the “Normalisation” project. The OCR’ing has been conducted with an expected accuracy of 90-95% [Sangiacomo et al. 2022b], sufficient for many text-mining methods, including for example vector-semantic modelling [Hill 2019] and collocate-extraction [Sangiacomo et al. 2022b].[12] 55

The texts are preprocessed by the lemmatization of the texts using the Natural Language Toolkit’s lemmatizer in Python [Bird et al. 2009], removing low frequency words (<4), removing numbers, interpunction and singular letters, decapitalization of all words, and the removal of OCR-based nonsense word types and articles. These steps are particular to this case-study, but the SDS-scores do not depend on any of these steps and can be applied to raw-texts as well as more heavily cleaned texts. 56

On the basis of this corpus we modeled 15 key terms’ SDS-scores. These key terms were derived both via domain knowledge and via topic modelling – a computational method that aims to extract the topics most central to the text.[13] That is to say, they play a central role within the corpus, but the list of these 15 words is in no sense exhaustive of such terms. 57

In Figure 2 we have provided the results of our explorative case-study as to the differences in stability one can find in salient and non-salient connections for technical terminology in early-modern English language natural philosophy/physics. For each term and for each work in the corpus, we have generated models based on the distributional algorithms mentioned above (one to be interpreted as a model of the salient elements of the terms’ meaning, and the other as the less salient parts of the terms’ meaning). Based on the outcomes, the terms will be classified in our system introduced in section 4. 58

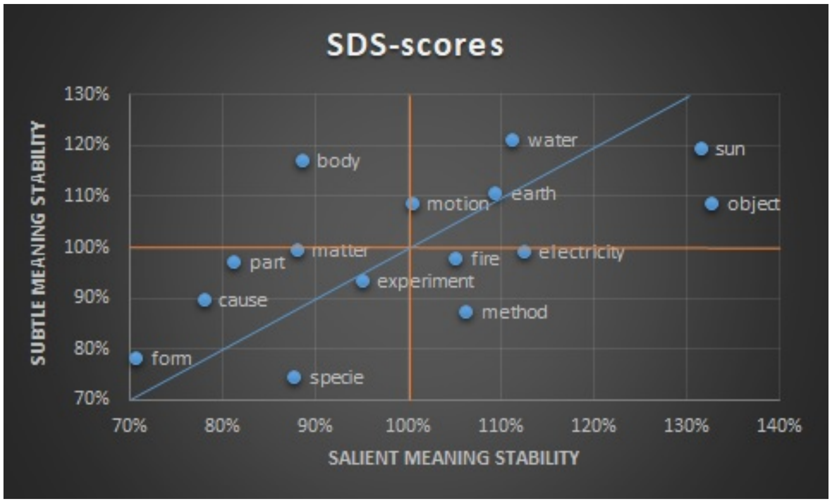

Figure 2. A scatter plot that shows positions of our investigated terms based on both their more and less salient stability scores.

Above we see a plot of the scores of our terminology. On the horizontal axis we have the average similarity normalized to the average corpus-wide similarity scores based on the collocation based approach, on the vertical, the same for the vector-based approach. From this, we are given four quadrants in which our values are plotted, which agree with the four categories we introduced in the previous section. The lower left quadrant signifies terms about which there is neither salient nor subtle stability in the terminology, the terms provoke discussion and their legitimacy is opposed. The upper right quadrant signifies terms that exhibit stability both in salient and subtle facets of their meaning – i.e. unproblematic, generally simple, terms. The lower right quadrant signifies terms that exhibit façade like behavior – the terms are stable and unproblematic at the surface level, but at the operative level exhibit unrecognized semantic shifts. Finally, the top left quadrant signifies what we have dubbed integral controversy terms – despite being central to much discussion, the terms exert a unifying power on the discourse through their subtle similarity. What follows is an interpretation of these results that places them within the literature on the development of early-modern natural philosophy – the case study is explorative and has as its main aim to show the possibility for fruitful application of the method. 59

Simple terms:

In the upper right quadrant we find 4 terms (excluding motion) – ‘earth’, ‘water’, ‘sun’ and ‘object’. The commonality between the first three terms is that they all refer to identifiable, concrete, phenomenally accessible, things. This is of course not to say there were no interesting discussions about these terms (a lot of astronomical work was done on the relation between the earth and the sun, and on the nature of these two planets) but the terms are still 60 relatively straightforward qua meaning. For the sun in particular a simple ostensive definition is available, somewhat similarly for water (water being this ostensively available stuff, and other similar stuff). Indeed, the three terms are all very neatly counted as easily identifiable, concrete, objects of investigation within the discipline of natural philosophy (fire and electricity also inch close to this quadrant). ‘Object’ is the only abstract term of the four and also scores somewhat closer to the façade quadrant than the rest. But, the usage of a term like ‘object’ is simple – and there can be little truly different applications of it. Because even when speaking of mental, biological or physical objects, in all cases nothing changes about the ways in which they are objects.

Integral controversy terms

There are three terms in our case-study that fall (or almost fall) in the upper left quadrant that we have proposed to understand as being comprised of terms that are extremely central to the discourse, are central to much controversy, but enforce a certain unity into the discourse via implicit agreement on the more general connections that such terms have. ‘Motion’ and ‘matter’ are edge cases, so let’s first look at the clearest case, ‘body’. 61

‘Body’ is extremely central to early modern natural philosophy. The early mechanicist philosophies in the 16th century (like Descartes’ philosophy) let themselves be summarized in the striking sentence “bodies in motion,” or, “matter in motion” [Nadler 1993, 3], [Roux 2017, 27–28]. Although many schools (already existing in the form of scholastic schools, or following schools, like experimentalists and Newtonians) will reject mechanicism’s claim that the investigation of “bodies in motion” exhausts the activities of the natural philosopher, they however will all have to concede that the mechanicist has set the program: to be a natural philosopher is to (at least) have answers to questions pertaining to bodies (whatever they might exactly be) and their movements (whatever that might end up meaning exactly). 62

Once the program has been set – the natural philosopher is to say something about bodies – it is natural that ‘body’ itself becomes a contested term. The philosopher who can provide and defend a conception of ‘body’ such that it easily fits with his/her more general outlook is a philosopher who has successfully scaled the walls of natural philosophy. At the same time, these moves need to be made out in the open – since the eye of the reader is fixed on these terms. Façade-like behavior is not to be expected – all of the disagreement is out in the open, not hidden, nor is there any manifest agreement that could help hide subterranean disagreement. 63

A few examples will show the extent of the mutability of ‘body’ within early modern philosophy, where these mutations are central to the arguments provided. Descartes’ mechanicism comes together with a very explicit statement about the nature of what bodies are – they are primarily to be characterized by their extension. This is coupled together with a statement about the other ‘type of thing’ there is in the world, namely, thought. Bodies being extension means that all the other properties of bodies can be explained reductively by reference to their extensional properties (and their motions) – except any properties (or movements) that are to be explained via thought. If body is anything, it is to be extension, and if body is to not be anything in particular, it is not to be thought. Amazingly, in opposition to this, we find later natural philosophers to remain true to Descartes style of reasoning about body (1. Body is central to their activity as natural philosophers and 2. It is an investigation into body’s essential attributes and motions that should occupy them in particular) while in the meantime moving away wholly from his conception of body. Indeed, to make room for their particular systems of natural philosophy, we find thinkers like Moore, Cavendish and Conway, more or less destroy all properties ascribed to bodies by Descartes. For Conway, both the impenetrability (one of the ways the space filling ‘extension’ was often fleshed out) of bodies is opposed and it is claimed that bodies are both spiritual and material-like [Lopston 1982, 15]. Cavendish phrases the issue somewhat differently, but for her too, bodies are coupled together with mental properties [Shaheen 2019, 3553–3554], as is the case for Moore, who arguably provides less reworking since he mostly takes the extended conception of motion to be gappy in accounting for phenomena [Roux 2017, 27]. What allows for these very deep metaphysical discussions is that, in the end, bodies can be somewhat easily identified in everyday activity. And whatever the exact nature of bodies turns out to be, it should in the end still be mapped on a set of paradigmatic instances of bodies (even the radical turnaround one finds in a thinker like Conway, where one would expect the concept to be so definitively deformed to no longer map unto the same objects as previous concepts of body, is still applied to recognizable cases, like the different unmixed fluids and how they hang together in a single bottle).[14] 64

So, what about ‘motion’ and ‘matter’? Generally, ‘matter’ was used more restrictively than ‘body’ – and was less easily identified with everyday objects in the way that bodies could be. One would expect that matter was less stable in application because matter was not so strictly tied together to phenomenally accessible paradigmatic examples (bodies might be best imagined as spheres, houses, coherent swathes of fluid, cannon balls whilst matter was often understood to encompass more restrictively the corpuscles that make up bodies). This is due to the Aristotelian roots of the concept – matter plays a role in the hylomorphic theories where things are essentially made up out of the combination of matter and form [Manning 2012]. 65

Motion get reinterpreted radically within early-modern philosophy, motion as change makes place for motion as mere local-motion. However, this move is not as extensively challenged in the development of early modern natural philosophy as was body. While body functioned as the pivotal point where different conceptualizations could be made to do all of the metaphysical lifting, motion did not take up such a place. Motion was on the agenda but was more easily understood and less prone to the radical reinterpretations (after the initial reconfiguration) than occurred for body. 66

Façade terms:

In the lower right quadrant we would expect to find terms that exhibit façade like behavior. These are words similar in salient ways but dissimilar in more subtle ways. Superficial similarity may mask slippery differences in usage. Three terms seem to apply – ‘fire’, ‘electricity’, and ‘method’. We will not discuss ‘fire’, because, as we can see in the plot the distance between ‘fire’ and the ‘x=x’ line is very small. That is to say, the difference between the more and less salient scores is very small (98/104). 67

Similarly, ‘electricity’ will not be extensively considered for two reasons. Firstly, it’s quite close on the line toward the upper right quadrant. In addition, the technical nature of the term precludes a useful analysis without more extensive background knowledge of the term. A few short remarks that should be read in that light are that: ‘electricity’ has properties in common with ‘earth’, ‘sun’ and ‘water’ in the sense that it is a term that designates a somewhat properly delineated group of phenomena. In particular, static electrical effects (which resulted in the attraction of other objects) and magnetic 68 materials (which show somewhat similar behavior) were termed electrical. Although practitioners disagreed on how to explain these phenomena, for some time, this set of phenomena was clearly delineated. However, further study revealed that some previously ‘electrical’ phenomena should be delineated from ‘the electrical per se’, in particular excluding magnetic phenomena from the term's extension [Gregory 2007, 35–42]. In addition, there is the novelty of the topic in natural philosophy as a systematic topic of interest – whereas it was of fringe importance in foregoing natural philosophy, it gets conceptualized more systematically within early modern natural philosophy – the novelty should induce us to expect less stability in the application of the term in its non-salient features, whereas identifiable phenomena should make us expect more stable non-salient applications. What is less easily explained is the cause for the agreement about salient facets of electricity’s meaning – especially since the debate about electricity is often characterized as a number of schools disagreeing fundamentally about the nature of electricity (Ibid, p.35-42). Note that most of the systematic early modern explanation started to come up in the second half of the 18th century, making it only a small part of the timeline of our corpus. We leave this discussion as is.

69 Central in the lower right quadrant is ‘method’. Many early modern thinkers agreed on the centrality of method in science and presumably its salient stability can be explained by this as well as by a shared understanding of the overall concept. We might be able to explain its subtle instability by the fact that method as a word generally signifies that some technicalities are to come, but which technicalities differs depending on which method in particular will be discussed.

70 If we are correct, method’s presence in this quadrant can be explained without calling it a façade. Electricity might be a façade, but to uncover this would require more in depth analysis of the development of the concept of electricity in this period. This means that our case-study did not bear out our expectations – no clear cases of facades were found in the lower right quadrant.

Crisis terms:

71 The most densely populated quadrant is that of the crisis terms (5 terms, excluding ‘matter’) – ‘part’, ‘cause’, ‘specie’, ‘form’, ‘experiment’. These are words which are understood differently and used differently. These terms are thus such that we should expect both that the terms were controversial and discussed explicitly in natural philosophy and that there were little methods available to tie these terms together via, for example, implicit agreement on the (paradigmatic) extension of these terms (as in the case of body).

72 All of these terms agree with this characterization. The four terms ‘part’, ‘cause’, ‘specie’ and ‘form’ are all derived from scholastic philosophy and heavily debated in early modern natural philosophy. Species, (substantial) forms and causes were all important aspects of the scholastic/Aristotelian framework, and were all reworked, or even outright rejected, in subsequent schools of natural philosophy (like Cartesianism and Newtonianism) [Blair 2006, 366]. Cause is already reworked by Descartes, who also rejects (although not fully) substantial forms [Flage 1997, 845]. We see species being rejected by mechanicists as well, as well as non-mechanicists like Conway [Conway 1996, 30–31]. Cause is outright rejected as being the proper object of investigation for natural philosophy by later Newtonians like van Musschenbroek [Sangiacomo 2018, 51]. Forms are sometimes reintroduced and reworked and all the while remain in play in the strong scholastic school that remains in operation for most of the early-modern period, particularly in the university context [Sangiacomo et al. 2022b]. What differentiates these terms from body are two things: i) body has easily accessible paradigmatic instances of the concept’s extension which are not so readily available for, for example, part and form (and arguably, at least some of the fourfold of Aristotelian causes) and ii) whereas body was extensively discussed in the light of its given central position to the discipline, these terms were discussed because their position within the discipline were under dispute. Experiment has a similar structure, except there the discussion will not have been most explicitly between scholastic and new philosophies, but between rationalist/Hobbesian conceptions of natural philosophy and experimental/Boylean [Shapin and Schaffer 1985/2018]. Again, not only is experiment a more technical term that allows for less easy identification of its extension (what even counts as an experiment, as opposed to observation, or an uninformative ‘account’) [Shapin and Schaffer 1985/2018]. In these debates, it is also still up for grabs whether it has any place in natural philosophy. In this sense we propose the terms in quadrant to be crisis terms – they are central to the discipline, because in many ways, the form of the discipline is transformed by drastically questioning the contents and validity of these terms.

Conclusion

73 In this paper two methods for the automated semantic investigations of corpora often used in the digital humanities have been compared over which facets of meaning they track. We’ve claimed that full vector-representations of a word’s meaning tracks the word’s more hidden, subtle and less salient aspects of its meaning and that the collocate-representation tracks the more salient aspects of its meaning. We’ve proposed a new measure on the basis of this which makes use of these features: Salience Differentiated Stability. SDS-scores signify the divergence between the salient stability and the non-salient of a word in a corpus of works. Finally, a small case study making use of SDS-scores concerning a corpus of early-modern natural philosophy has been provided.

Acknowledgements

74 This research has been supported by the ERC starting grant NaturalPhilosophy Grant ID: 801653, with the title: “The normalization of natural philosophy: how teaching practices shaped the evolution of early modern science” and the Research Training Group Situated Cognition funded by the Deutsche Forschungsgemeinschaft .

Addendum

75 This addendum contains some details about the contents of our corpus – all of the following texts have been made use of in OCR’d format to allow for the computational analyses given above.

Author	Title	Publication

		Date
Digby	Two Treatises in the one of which, The nature of Bodies; in the other, The nature of Mans Soule, is looked into	1644
White	Institutiones Peripateticae	1646
Comenius	Naturall philosophie reformed by divine light, or, A synopsis of physicks by J.A. Comenius	1651
Charleton	Physiologia Epicuro-Gassendo-Charltoniana: or a Fabrik of Science Natural, Upon the Hypothesis of Atoms, Founded by Epicurus, Reparied by Petrus Gassendus, Augmented by Walter Charleton	1654
Boyle	The Sceptical Chymist (2 ed: ... Whereunto is Added a Defence of the Authors Explication of the Experiments, Against the Obiections of Franciscus Linus and Thomas Hobbes (a book-length addendum to the second edition of New Experiments Physico-Mechanical)	1661
Boyle	Considerations touching the Usefulness of Experimental Natural Philosophy (followed by a second part in 1671)	1663
Boyle	Origin of Forms and Qualities according to the Corpuscular Philosophy	1666
Cavendish	Observations on Experimental Philosophy	1666
Cavendish	Grounds of Natural Philosophy	1668
Boyle	A Free Enquiry into the Vulgarly Received Notion of Nature	1686
Midgley	A new treatise of natural philosophy, free'd from the intricacies of the schools adorned with many curious experiments both medicinal and chymical : as also with several observations useful for the health of the body	1687
Le Grand	An Entire Body of Philosophy according to the principles of the famous Renate Des Cartes in three books.	1694
Newton	Opticks	1704
Ditton	The General Laws of Nature and Motion, with their application to mechanicks. Also the doctrine of centripetal forces and velocities of bodies describing any of the conick sections, being a part of the great Mr. Newton's principles	1709
Worster	A compendious and methodical account of the principles of natural philosophy	1722
Stirling	A course of mechanical and experimental philosophy : consisting of Seven Parts.	1727
Greene	The Principles of the Philosophy of the Expansive and Contractive Forces; or, An Inquiry into the Principles of the Modern Philosophy, that is, into the Several Chief Rational Sciences, which are Extant	1727
Pemberton	A View of Sir Isaac Newton's Philosophy	1728
Horsley	A Short and General Account of the most Necessary and Fundamental Principles of Natural Philosophy ... Revised, corrected, and adapted to a course of experiments ... By John Booth.	1743
Desaguliers	A course of Experimental Philosophy	1745
Rowning	A Compendious System of Natural Philosophy	1744
Robertson	The principles of natural philosophy explain'd and illustrated by experiments; in a course of sixteen lectures: to be perform'd at Mr. Fuller's Academy, etc.	1745
Knight	An attempt to demonstrate that all the phaenomena in nature may be explained by two simple active principles, Attraction and Repulsion: Wherein the Attraction of Cohesion, Gravity and Magnetism are shewn to be One and the Same and the Phenomena of the Latter are more Particularly Explained	1748
Rutherfort	A System of Natural Philosophy; being a Course of Lectures in Mechanics, Optics, Hydrostatics and Astronomy	1748
MacLaurin	An Account of Sir Isaac Newton's Philosophyical Discoveries	1748
Emerson	The Principles of Mechanics: explaining and demonstrating the general laws of motion, the laws of motion, the laws of gravity, motion of descending Bodies, projectiles, mechanics powers, pendulums, center of gravity etc. strength and stress of timber, hydrostatics, and construction of machines	1754
Wilson	The principles of philosophy. The principles of natural philosophy: with some remarks upon the fundamental principles of the Newtonian philosophy	1754
Jones	An essay on the first principles of natural philosophy ... in four books	1763
Ferguson	An easy and pleasant introduction to Sir Isaac Newton's Philosophy : containing the first principles of mechanics, trigonometry, optics, and astronomy	1772
Arden	Analysis of Mr. Arden's course of lectures on natural and experimental philosophy. Viz. Natural philosophy in general, chemistry, electricity, mechanics, geography, astronomy, hydrostatics, pneumatics, optics	1774
Hamilton	Four introductory lectures in natural philosophy. : I. Of the rules of philosophising, the essential properties of matter, and laws of motion. II. Of the several kinds of attraction, and particularly of cohesion. III. Of gravity, or the attraction of gravitation. IV. The laws of motion explained, and confirmed by experiments.	1774
Lovett	The Electrical philosophers, Containing a New System of Physics	1774
Fenning	The young man's book of knowledge : being a proper supplement to The young man's companion. In six parts	1774
Banks	An epitome of a course of lectures on natural and experimental philosophy.	1775
Atwood	A Description of the Experiments intended to Illustrate a Course of Lectures on the Principles of Natural philosophy	1776
Cullen	First lines of physics, for the use of students in the University of Edinburgh (4 vol.)	1777
Robinson	Outlines of a Course of Experimental Philosophy	1784
Atkinson	A Compendium of a Course of Lectures on Natural and Experimental Philosophy	1784
Telescope	The Newtonian system of philosophy.	1787

Peart	On the elementary principles of nature ; and the simple laws by which they are governed. Being an attempt to demonstrate their existence, and to explain their mode of action ; particularly in those states, in which they produce the attractions of cohesion, gravitation, magnetism and electricity ; and also fire, light, and water.	1789
Adams	Lectures on natural and experimental philosophy, considered in it's present state of improvement.	1794
Walker	Analysis of a course of lectures in natural and experimental philosophy : Viz. 1. Properties of matter, 2. Mechanics, 3. Chemistry, 4 & 5. Pneumatics, 6. Hydrostatics, 7. Electricity, 8. Electricity, 9. Optics, 10. Use Of The Globes, 11 & 12. Astronomy, &c. By A. Walker.	1795
Anderson	Institutes of Physics	1798
Gregory	The Economy of Nature Explained and Illustrated on the Principles of Modern Philosophy vol.2	1798
Gregory	The Economy of Nature Explained and Illustrated on the Principles of Modern Philosophy vol.3	1798
Wood	The principles of mechanics : designed for the use of students in the University. By James Wood ...	1800
Robison	Elements of Mechanical Philosophy: Being the Substance of a Course of Lectures on that Science	1804
Young	A course of lectures on Natural Philosophy and the Mechanical Arts	1807
Playfair	Outlines of Natural Philosophy	1812
Wylde	The circle of the sciences; a cyclopaedia of experimental, chemical, mathematical, & mechanical philosophy, and natural history;	1862

Table 2. Corpus

Notes

[1] For a good general introduction, see [Smith 2020]

[2] For example, the problem with a very small windowsize is that it starts picking up on syntactic relations moreso than semantic relations (verbs will generally be scored lower with one another due to the syntactic restrictions on following up verbs with verbs for example) – meaning windowsize is not just an optimizable parameter, but one that can be tweaked to pick up on different features that may prove to be of interest. We have followed a windowsize that (very roughly) is small enough to approximate a ‘sentence length’ window, while being large enough to smooth out some of the syntactical connections – but the choice for 10 and not, say, 15, is somewhat arbitrarily made here. Other variations are also available, see for example [de Bolla 2019, 375–378]

[3] Introduced for applications in linguistic analysis by Church and Hanks in *Word Association Norms, Mutual Information, and Lexicography* (1989). Used and discussed further in for example [Bouma 2009].

[4] This is analogous to methods of finding out how two subsets of a probability space are correlated. There the normal procedure is to divide the chance of being in both by the chance of being in both under the hypothesis that they are independent: $P(x, y)/P(x)P(y)$. If the result is 1, they are independent, if higher, the two sets are positively correlated with higher numbers indicating stronger correlations, and if lower, they are anticorrelated, with lower numbers indicating stronger anticorrelation. Here we use the same idea albeit slightly complicated by the fact that we are working with windows around tokens of x, and that instances of y may be counted doubly. In most formulations of PMI [Church 1989], the notation mimics more closely finding correlations between subsets of a probability space – computationally, we do the same thing as these authors.

[5] For examples and applications of this sort of analysis, see [Gavin 2019], [Brezina 2015], [Davies 2012] and, for a particularly influential application, see [de Bolla 2013]

[6] The choice for the minimal threshold for the PMI value might appear arbitrary, while also being significantly influential in the differences between the output of the two methods we propose to give an interpretation. However, as we argued in the introduction, although there is here a paradox of the heap, at least an interpretation is available for the quantity (we propose salience) that is being cut-off.

[7] Especially for historical analysis of texts, these methods are more useful than the often more complicated, fully quantitative and hard to interpret tools of natural language processing.

[8] The suggestion for the application of these methods can already be found in [Harris 1964, 158–161]

[9] As an example see [Underwood 2019]

[10] [Brezina 2015] have provided a good framework for combining vector semantics and network analysis in more involved ways than those that we have applied here.

[11] See “Mapping early modern natural philosophy: corpus collection and authority acknowledgement” [Sangiacomo et al. 2021b]. For more information on the construction process. And for more information on the relative merits compared to classical corpus construction, see “Expanding the Corpus of Early Modern Natural Philosophy: Initial Results and a Review of Available Sources” [Sangiacomo et al. 2021a].

[12] Note that the corpus has been built up without preference to particular philosophical schools, and contains a spectrum of natural philosophical schools and their works. The titles, authors and dates of our sub corpus can be found in the addendum, a more complete overview of the entire corpus can be found online [Sangiacomo et al. 2021b].

[13] For the topic modelling in particular we thank Raluca Tanasescu who modelled part of a similar, Latin corpus, from this we have taken suitable translations.

[14] She argues that a phenomenon like the freezing of a bottle of liquor producing a small unfrozen part with a higher density of alcohol (or, tentatively, spirit) can best be modeled as a separation of gross ‘body’ from the more supple ‘spiritual’ body (i.e. the unmixing of a liquid through freezing is a separation of the more spiritual bodies from the more matter-like ones.) [Conway 1996, 43–44]

Works Cited

Bird et al. 2009 Bird, Steven, Loper, Edward and Klein, Ewan. (2009), *Natural Language Processing with Python*. O’Reilly Media Inc.

Blair 2006 Blair, A. (2006). “Natural Philosophy.” In K. Park, & L. Daston, *The Cambridge History of Science; Early Modern Science* (Vol. 3, pp. 365-405). Cambridge: Cambridge University Press.

Bouma 2009 Bouma, G. (2009). “Normalized (pointwise) mutual information in collocation extraction.” *Proceeding of GSCL*, (pp. 31-40).

Brezina 2015 Brezina, V., McEnery, T., & Wattam, S. (2015). “Collocations in context: A new perspective on collocation networks.” *International Journal of Corpus Linguistics, 20*(2), 139-173.

Church 1989 Church, K., & Hanks, P. (1989). “Word Association Norms, Mutual Information, and lexicography.” *27th Annual Meeting of the Association for Computational Linguistics*, 76-83.

Conway 1996 Conway, A. (1996). *The Principles of the Most Ancient and Modern Philosophy.* (A. Coudert, & T. Corse, Eds.) Cambridge: Cambridge University Press.

Davies 2012 Davies, M. (2012). “Expanding horizons in historical linguistics with the 400-million word Corpus of Historical American English.” *Corpora,* 7, 121-157

Devilin 2018 Devlin, J. Chang, M. Lee, K. & Toutanova, K. (2018). “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.” *arXiv:1810.04805v2*

Firth 1957 Firth, J. (1957). *A Synopsis of Linguistic Theory.* Oxford: *Studies in linguistic analysis*, Blackwell.

Flage 1997 Flage, D., & Bonnen, C. (1997). “Descartes on Causation.” *The Review of Metaphysics, 50*(4), 841-872.

Gal 2013 Gal, O., & Chen-Morris, R. (2013). *Baroque Science.* Chicago: The University of Chicago Press.

Gavin 2019 Gavin, Jennings, Kersey, & Pasanek. (2019). Spaces of Meaning: Conceptual History, Vector Semantics, and Close Reading. In Gold, & Klein, *Debates in the Digital Humanities 2019.* Minneapolis: University of Minnesota Press.

Goldberg 2014 Goldberg, Y. & Levy, O. (2014) “word2vec Explained: Deriving Mikolov et al.’s Negative-Sampling Word-Embedding Method.” *arXiv:1402.3722*

Gregory 2007 Gregory, F. (2007). *Natural Science in Western History.* Boston: Houghton Miflin Company.

Hamilton 2016 Hamilton, W. Leskovec, J. Jurafsky, D. (2016) “Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change.” arXiv:1605.09096

Han 2012 Han, J., Kamber, M., & Pei, J. (2012). “Getting to Know Your Data.” In *Data Mining* (Third edition ed., pp. 39-82). Elsevier.

Harris 1964 Harris, Z. (1964). “Distributional Structure (org. 1954).” In J. Fodor, & J. Katz, *The Structure of Language: Readings in the Philosophy of Language* (pp. 33–49). Englewood Cliffs: Prentice-Hall.

Haslanger 2012 Haslanger, S. (2012). “What Are We Talking About? The Semantics and Politics of Social Kinds.” In S. Haslanger, *Resisting Reality: Social Construction and Social Critique* (pp. 365-380). Oxford University Press.

Hill 2019 Hill, M.J., Hengchen, S. (2019) “Quantifying the impact of dirty OCR on historical text analysis: Eighteenth Century Collections Online as a case study, Digital Scholarship in the Humanities,” 34 (4): 825–843, https://doi.org/10.1093/llc/fqz024.

Hughes et al. 2016 Hughes, L., Constantopulos, P., & Dallas, C. (2016). “Digital Methods in the Humanities: Understanding and Describing their Use across the Disciplines.” In R. S. Susan Schreibman, *A New Companion to Digital Humanities* (pp. 150-170). Oxford, Malden: John Wiley & Sons, Ltd.

Kuhn 2012 Kuhn, T. (1962/2012). *The Structure of Scientific Revolutions* (4th ed.). Chicago: University of Chicago Press.

Landauer and Dumais 1997 Landauer, T., & Dumais, S. (1997). “A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge.” *Psychological review, 104*(2), 211.

Lopston 1982 Lopston, P. (1982). “Introduction.” In A. Conway, & P. Lopston (Ed.), *The Principles of the Most Ancient and Modern Philosophy* (pp. 1-60). Dordrecht: Kluwer Academic Publishing Group.

Lowe and McDonald 2000 Lowe, W., & McDonald, S. (2000). *The direct route: Mediated priming in semantic space.* The University of Edinburgh.

Ludlow 2014 Ludlow, P. (2014). *Living Words: Meaning Underdetermination and the Dynamic Lexicon.* Oxford University Press.

Manning 2012 Manning, G. (2012). “Three Biased Reminders about Hylomorphism in Early Modern Science and Philosophy.” In *Matter and Form in Early Modern Science and Philosophy* (pp. 1-32). Leiden: Brill.

McDonald 2000 McDonald, S. (2000). *Environmental Determinants of Lexical Processing Effort* (PhD-Thesis ed.). University of Edinburgh.

McDonald and Ramscar 2001 McDonald, S., & Ramscar, M. (2001). “Testing the distributioanl hypothesis: The influence of context on judgements of semantic similarity.” *Proceeding of the Annual Meeting of the Cognitive Science Society*, *23.*

Mikolov et al. 2013 Mikolov, T. Chen, K. Corrado, G. & Dean, J. (2013). “Efficient Estimation of Word Representations in Vector Space.” *arXiv:1301.3781v3*

Nadler 1993 Nadler, S. (1993). *Introduction.* In *Causation in early modern philosophy* (pp. 1-8). University Park: The Pennsylvania State University Press.

Ramscar and Yarlett 2000 Ramscar, M., & Yarlett, D. (2000). “The use of a high-dimensional, 'environmental' context space to model retrieval in analogy and similarity-based transfer.” *Proceedings of the 22nd Annual Conference of the Cognitive Science Society.*

Roux 2017 Roux, S. (2017). “From the mechanical philosophy to early modern mechanisms.” In P. I. Stuart Glennan, *The Routledge Handbook of Mechanisms and Mechanical Philosophy* (pp. 26-45). Routledge.

Sangiacomo 2018 Sangiacomo, A. (2018). “Teleology and the Evolution of Natural Philosophy.” *Studia Leibnitiana, 50*(1), 41-56.

Sangiacomo et al. 2021a Sangiacomo, A., Tanasescu, R., Donker, S., & Hogenbirk, H (2021a). Expanding the Corpus of Early Modern Natural Philosophy: Initial Results and a Review of Available Sources. Journal of Early Modern Studies, 10(1), 107-115

Sangiacomo et al. 2021b Sangiacomo, Andrea; Tanasescu, Raluca; Donker, Silvia; and Hogenbirk, Hugo (2021b). “Normalisation of Early Modern Science: Inventory of 17th- and 18th-Century Sources (1.0.0”) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.5566681

Sangiacomo et al. 2022a Sangiacomo, A., Tanasescu, R., Donker, S., & Hogenbirk, H. (2022a). “Mapping early modern natural philosophy: corpus collection and authority acknowledgement.” *Annals of Science, 79(1) 1-39*.

Sangiacomo et al. 2022b Sangiacomo, A., Hogenbirk, H., Tanasescu, R., Karaisl, A., White, N. (2022b) “Reading in the mist: High-quality optical character recognition based on freely available early modern digitized books.” *Digital Scholarship in the Humanities*.

Shaheen 2019 Shaheen, J. (2019). *Part of nature and division in Margaret Cavendish’s materialism. Synthese, 196*, 3551-3575.

Shapin and Schaffer 1985/2018 Shapin, S., & Schaffer, S. (1985/2018). *Leviathan and the air-pump* (First Princeton Classics paperback ed.). Linotron Baskerville: Princeton University Press.

Smith 2020 Smith, N. A. (2020). *Contextual Word Representations: A Contextual Introduction.* Retrieved from https://arxiv.org/: https://arxiv.org/pdf/1902.06006.pdf

Underwood 2019 Underwood, T. (2019). *Distant Horizons: Digital Evidence and Literary Change.* Chicago: University of Chicago Press.

Wevers and Koolen 2020 Wevers, M. & Koolen, M. (2020). Digital begriffsgeschichte: Tracing semantic change using word embeddings. *Historical Methods: A Journal of Quantitative and Interdisciplinary History,* 53, 226-243

Wilson 2006 Wilson, M. (2006). *Wandering Significance: An Essay on Conceptual Behavior.* Oxford: Clarendon Press.

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

de Bolla 2019 de Bolla, P., Jones, E., Nulty, P., Recchia, G., & Regan, J. (2019). “Distributional Concept Analysis: A Computational Model for History of Concepts.” *Contributions to the history of concepts*, 66-92.