



This is a repository copy of *Volatile concepts*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/194380/>

Version: Published Version

---

**Article:**

Fitzmaurice, S. [orcid.org/0000-0002-8804-1987](https://orcid.org/0000-0002-8804-1987) and Mehl, S. [orcid.org/0000-0003-3036-8132](https://orcid.org/0000-0003-3036-8132) (2022) *Volatile concepts*. *Corpus studies of language through time*, 27 (4). pp. 428-450. ISSN 1384-6655

<https://doi.org/10.1075/ijcl.22005.fit>

---

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>

# Volatile concepts

## Analysing discursive change through underspecification in co-occurrence quads

Susan Fitzmaurice and Seth Mehl  
University of Sheffield

This paper demonstrates the value of studying co-occurrence ‘quads’ – constellations of four non-adjacent lemmas that consistently co-occur across spans of up to 100 tokens – for understanding discursive change. We map meaning onto quads as ‘discursive concepts’, which encompass encyclopaedic semantics, pragmatics, and context. We investigate a high-frequency quad with high co-occurrence strength in EEBO-TCP: *world-heaven-earth-power*. We conduct semantic and pragmatic analysis to generate hypotheses regarding discursive change. The quad’s components are semantically underspecified; thus, although the quad indicates a discursive concept, each instantiation of the quad is variable, contingent, and dependent upon context and pragmatic processes for interpretation. We observe how the vague lexemes that constitute building blocks of religious discourse are employed to generate new, timely secular discourses; and we argue that semantic underspecification is the site and source of discursive change. Indeed, the volatile, unstable nature of the component lexical meanings renders them indispensable to early modern debate.

**Keywords:** co-occurrence, semantics, pragmatics, discourse, Early Modern English

### 1. Introduction

This study investigates discursive ‘quads’ – co-occurrences of four non-adjacent lemmas within large, discursive proximity windows – identified computationally across very large text collections. The identification and analysis of quads is the aim of a new computational process called ‘linguistic concept modelling’ (see Mehl, forthcoming; ‘Methodology’ below), which developed from The Linguistic DNA project (AH/M00614X/1). In this study, we employ linguistic concept mod-

elling to identify and quantitatively analyse quads in Early English Books Online (Text Creation Partnership edition; EEBO-TCP), which contains approximately 60,000 texts, and over 1.2 billion words, printed in English or in the British Isles between 1476 and 1700. We then conduct close, manual semantic and pragmatic analysis of one high-frequency quad, *world-heaven-earth-power*, based on the methodology demonstrated by Fitzmaurice (forthcoming), mapping the quad's variation and change over time, alongside variation and change to the discourses it contributes to. Discourse in this discussion refers to the mode of organizing knowledge, ideas or experience, together with the social practices, forms of subjectivity and power relations, which relates to a particular domain of intellectual or social activity, for example, political and religious discourses. Discourse consists of language, but is interpreted as “practices that systematically form the objects of which they speak” (Foucault, 2003: 49). The quad's component lemmas are themselves high-frequency, and also highly polysemous and vague. We demonstrate that such co-occurrences, rather than mutually disambiguating or clarifying the lexical meaning of the component parts, actually result in aggregated underspecification (vagueness), and we argue that this aggregated underspecification is heightened as we move from co-occurrence pairs to larger co-occurrence groupings such as quads. Our analysis leads iteratively to our hypothesis that the vague lexemes that are the building blocks of strictly religious discourse are recycled as a salient backdrop for newly secular discourse; and illustrates our argument that semantic underspecification is the site and source of discursive change. Indeed, the volatile, unstable nature of the quad's component lexical meanings, and the active work required for interpretation, render these concepts basic and indispensable to early modern debate. Below, we first present our perspective on aggregated underspecification. We then describe our data and methods and present our analysis of *earth-heaven-world-power* in EEBO-TCP.

## 2. Aggregated underspecification

It is an established practice in corpus semantics to study meanings and uses of individual lexical items via their co-occurrence pairs; a list of co-occurrence pairs informs an analysis of a selected individual node word: the meaning of the node varies and changes, and the changing pairwise co-occurrences evidence that variation and change (Sinclair, 1991). For example, the co-occurrence of *poor* with *beggar* and *rob(bing)* decreases during the 17th century, while co-occurrences with *sighing* and *helpless* increase, indicating that *poor* increasingly evoked empathy rather than fear for 17th century writers and readers (McEnery & Baker, 2017).

Less studied is the fact that the meanings of each item in the co-occurrence pair can vary and change simultaneously and in parallel, thus complicating interpretation. Meanings in pairs can vary and change in relation to polysemy or vagueness in either term (on polysemy and vagueness, see Geeraerts, 1993/2006; Cruse, 2011). In the case of polysemy, a pair's meaning can change discretely. For example, the pair *air-organ* could occur in a medical text about breathing, or in a musical text about a melody played on an instrument, because *air* is discretely polysemous (with the meanings "atmospheric gas" and "melody"; OED, s.v. *air* (I, II)), as is *organ* (with the meanings "part of the body" and "musical instrument"; OED, s.v. *organ* (I, IV)). The variation and change caused by discrete polysemy are easy to discern by reading examples in context, even if they can cause problems in interpreting large-scale calculations of co-occurrence data in big text collections.

In the case of vagueness, a pair's meaning can shift because of underspecification: for example, *security-feeling* is underspecified in multiple ways. *Security* is underspecified insofar as it can refer to the stability, safety, or well-being of any number of entities: an individual or a group (going back to the 15th century; OED, s.v. *security* (I.2.a)); or a nation or state (going back to the 17th century; OED, s.v. *security* (I.2.b)); or various levels of organisation between the individual and the state, which might be powerful (like the state) or disempowered, including families or communities. *Feeling* is underspecified insofar as it can indicate the capacity to experience sentiment, emotion, or opinion (OED, s.v. *feeling* (II.3)); the state of being affected by an emotion, sentiment, or opinion (OED, s.v. *feeling* (II.4.a)); the emotion, sentiment, or opinion itself (OED, s.v. *feeling* (II.4.b)); or indeed any point along a spectrum between those meanings (among other meanings). Because of these underspecifications, an observation that *feeling* appears near *security* provides little clear additional information about the meaning of *security*, and vice versa. That is, these underspecifications, instead of disambiguating or clarifying, can cumulatively aggregate: the meaning of the co-occurrence pair *security-feeling* is even more complex than that of either *security* or *feeling*, and the multiple interpretations are not discrete like the multiple interpretations of *air-organ*. The underspecification emerges from – and in spite of – the meaning contributions of each lexical item. Co-occurrence examples of *security-feeling* can evoke different points along a spectrum for each word's underspecified meanings, in ways that might change over time, and might consistently vary across discourses, text types, genres, or populations of speakers or writers – or even across multiple uses by a single writer, in a single passage of text.

The Linguistic DNA project (AH/Moo614X/1) recognised these issues as the project mapped prominent lexical co-occurrences in EEBO-TCP. The issues were magnified because the project examined not only co-occurrence pairs, but also

quads (Mehl, forthcoming); and developed a focus on highly frequent lemmas, which also tend to be highly vague, i.e. underspecified (Mehl, forthcoming). Moving from pairs to highly vague trios and quads, underspecification aggregates further, complicating interpretation even more. For example, *family* has undergone semantic change related to shifting social and cultural facts about households (Tadmor, 1996). Using established methods for studying semantic and conceptual change via lexical co-occurrence (see McEnery & Baker, 2017), we would expect to see changing co-occurrence pairs with *family* over time, and we might interpret these as clear and stable facts about semantic change. With our focus here, we might instead investigate how extremely complex discourses are built up around malleable co-occurrences beyond the level of the pair, like *family-home-control*, *family-marriage-blood*, or *family-lineage-society*, whose lexical items are themselves extremely complex and underspecified, with dramatic implications for meaning and discourse in relation to society and culture. The meanings of co-occurrences like these can be observed to vary and change in and around their underspecificity – and because of their underspecificity. It is precisely the accumulated underspecification within highly frequent co-occurrence quads that renders them invaluable to discourse. They are profoundly influenced by shifting social and cultural contexts – the sorts of contexts that provide the necessary enrichment to underspecified terms (Fitzmaurice, forthcoming) – and they are intensely susceptible to redefinition and manipulation to reflect changing social and cultural needs. They are therefore uniquely powerful building blocks of social and cultural discourse and debate.

One consequence of our approach is that the implicit focus in corpus linguistics on co-occurrences as stable indicators of meaning ought to be adjusted. While archetypal, clear, stable examples like *strong tea* and *powerful engine* are valuable, we see a significant field of study opening up which investigates how co-occurrence pairs, trios, or quads containing highly frequent, highly underspecified lexical items might vary and change across populations and time, and how that variation and change relate to the negotiation of underspecificity in social and cultural contexts. Indeed, *power* is underspecified in important ways (as we demonstrate below), and even *engine* is both polysemous and vague diachronically, such that an iconic example like *powerful engine* is a pair whose meaning could change dramatically over the centuries.

### 3. Methodology

The data for this study, EEBO-TCP, can be seen as an arbitrary collection (see Fitzmaurice et al., 2017), rather than a systematic sample of a well-defined popu-

lation, and it would therefore be inappropriate to generalise results to language, domains or discourses beyond the corpus. EEBO-TCP nonetheless facilitates valuable access to very large historical text data, and allows for reproducible experiments (McEnery & Baker, 2017: 55). Our version of EEBO-TCP is part-of-speech tagged and lemmatised using *MorphAdorner* (Burns, 2013), with additional automated and manual correction by the *MorphAdorner* team (M. Mueller, personal communication, 2018); it is not yet publicly available.

Linguistic concept modelling (Mehl, forthcoming) was applied to EEBO-TCP. Linguistic concept modelling was designed to provide a new way in to studying conceptual and discursive features of very large text collections by identifying concepts and discourse that might never have been identified using other approaches. The method computationally identifies constellations of lemmas that co-occur beyond the level of the phrase or clause, at the level of discourse, i.e. within 101-token spans (50 tokens to either side of a node). This sort of co-occurrence in wide proximity windows is comparable to Hoey's (2004, 2005) 'text collocation'. Like text collocation, and unlike *n*-grams, our co-occurrences may be non-adjacent, occurring anywhere across the wide proximity window. Instead of identifying pairs, our algorithms identify trios and quads, by first identifying a node, and then counting second, third, and fourth lemma co-occurrences within a window of 50 tokens to the left and right of the node. The process is computationally expensive, requiring High Performance Computing; it employs the big data software utility *Hadoop* and processes data across up to 40 virtual machines. Early experiments were conducted using proximity windows of 5, 10, 30, 50, and 100 tokens to the left and right of the node (see Mehl, 2019): proximity windows of 5 or 10 tokens tend to identify lexical relationships at the level of the phrase or clause, which was deemed unsuitable for our interest in discourse-level concepts, while a proximity window of 100 tokens demands too much processing time and storage space; a proximity window of 50 was deemed a reasonably large discursive span that is not prohibitively demanding computationally (while a proximity window of 30 remains a viable alternative, which deserves further experimentation). While trios are less computationally demanding than quads, our team has found that quads provide much richer semantic, pragmatic, and discursive material for analysis; 'quintos' remain prohibitively computationally expensive.

For each quad, our algorithms calculate raw frequency, chi-square score, and a modified MI score based on Fano (1960), which incorporates a novel part-of-speech baseline (see Mehl, 2019, for details of the statistical analysis). We then calculate 5 'bands' for frequency and MI, respectively (see [www.linguisticdna.org/cmd](http://www.linguisticdna.org/cmd) for details on calculating bands). Ranking and sorting is thus possible via multiple combinations of raw frequency, MI score, and frequency and MI bands. Quads that do not pass a chi-square threshold ( $p < 0.05$ ) are excluded; combining

chi-square and MI in this way is standard in corpus linguistics and, combined with frequency thresholds (described below), mitigates the well-known ‘weighting’ of low-frequency items in MI scores (see Mehl, 2019).

The Linguistic DNA project has identified billions of quads in EEBO-TCP, with associated statistical scores. In various iterations of our algorithms, we have analysed trios or quads around nodes and co-occurrences that occur between 2 and 5,000 times each in EEBO-TCP, with various restrictions according to parts of speech and chi-square threshold: stricter thresholds reduce computational demands. Our current iteration identifies node and co-occurring lemmas that each occur at least 5,000 times in EEBO-TCP; in which each component is a noun, adjective, or verb; and which are statistically significant at  $p < 0.05$ . The current public interface for our data presents quads according to those thresholds, built around 75 of the highest frequency nouns in EEBO-TCP ([www.linguisticdna.org](http://www.linguisticdna.org)).

Linguistic concept modelling is an exploratory tool, valuable for generating hypotheses, which can reveal associations that would not be identifiable otherwise. This paper is thus a case study for this exploratory process, in which computational processes and statistical analysis open a door to principled, qualitative semantic and pragmatic analysis on a single quad across a large text collection. The quad investigated here is just one among at least thousands that deserve further attention, both individually and in comparison to each other, and we invite other scholars to conduct such investigations with the Linguistic DNA concept modelling demonstrator ([www.linguisticdna.org](http://www.linguisticdna.org)).

The quad *power-heaven-earth-world* is among the most frequently occurring quads in EEBO-TCP, occurring 12,187 times, with an MI score of 4.6. For comparison, we present below the five highest-frequency quads around the high-frequency lemmas *earth* (Table 1) and *power* (Table 2) respectively, with MI scores in bands 3 to 5 (i.e. relatively high MI), with the thresholds and parameters outlined above.

**Table 1.** High frequency quads around node word *earth*

Quad	Frequency in EEBO-TCP	MI score
<i>earth-air-fire-water</i>	32,759	7.2
<i>earth-day-heaven-lord</i>	22,364	4.2
<i>earth-great-heaven-lord</i>	21,535	3.9
<i>earth-heaven-new-see</i>	20,093	5.6
<i>earth-glory-heaven-lord</i>	19,514	3.9

**Table 2.** High frequency quads around node word *power*

Quad	Frequency in EEBO-TCP	MI score
<i>power-king-law-people</i>	35,149	5.6
<i>power-king-law-parliament</i>	34,198	7.4
<i>power-father-holy-son</i>	23,192	6.1
<i>power-contain-line-superficies</i>	22,362	17.0
<i>power-father-ghost-holy</i>	21,896	5.8

Crucially for this study, the component lemmas of *power-heaven-earth-world*, in addition to being high-frequency in EEBO-TCP, are highly polysemous and vague, as reflected in OED entries for each, and as elaborated in the study below. Finally, the quad contains lemmas whose underspecificity spreads across religious and non-religious domains. For example, *power* is often attributed either to God or to royalty, or indeed to lay readers of the Bible, or to others, sometimes in combination. *World* can refer to the whole of human and divine existence, or to the realm of human experience alone, the latter occurring with or without connotations of sinfulness and corruption. This quad therefore provides a suitable way in to examining the value of underspecificity in discourse and debate that may be politically or religiously charged, and the ways that underspecificity relates to change over time.

To analyse individual and collective examples of each quad, we proceed in stepwise fashion through a close pragmatic analysis, as defined and illustrated by Fitzmaurice (forthcoming), for identifying, dissecting, and synthesizing ‘discursive concepts’ indicated by the quads. We define ‘discursive concepts’ as encompassing encyclopaedic semantics, pragmatics, and context, and not co-terminous with any of the lexical components of the quad (see Fitzmaurice et al., 2017). We call this stepwise process ‘discursive concept analysis’. These steps begin with an analysis of each lemma, and proceed through the inspection of the discursive relationship between the lemmas, through the narrow co-text, the wider co-text, and then broader textual and historical contexts of use. For each example of the quad, this entails:

- i. the semasiological inspection of each lemma in use, to disambiguate polysemy and clarify underspecification by ascertaining the most likely dictionary sense or sub-sense depending upon the immediate co-text, and identifying underspecified semantic features that could allow for pragmatic enrichment for specification – for which we refer extensively to the OED;



- ii. the 'resolution of reference' by 'assigning an appropriate contextual value' to each lemma in relation to the others, including identifying discourse prosody and other evaluative information in narrow co-text;
- iii. the assignment of pragmatic meaning to the quad, informed by the wider co-text and the broader social and cultural context, including narrowing, broadening, and specifying of the discourse itself, based on the contributions of each lemma (with its associated underspecification), the lemmas collectively (with their aggregated underspecification), co-text, and context;
- iv. the description of a 'discursive concept' indicated by the quad, which includes semantics, pragmatics, and discourse, and is not co-terminous with the words that constitute the quad.

These processes are necessarily iterative; the construction of a discursive relationship between two members of a quad has the potential to destabilise the relationship inferred between that pair and the third member of a quad, or between a trio and the fourth member of the quad. That is, the stepwise analysis of the nature of the relationship between highly underspecified lemmas involves testing the interpretation formed in light of any additional information invoked or presented through the analysis of other members of the quad. This interpretation is assessed and adjusted with the addition to the analysis of each additional lemma, thus destabilising the tentative conclusions we might have drawn about the relationships among lemmas we have already analysed. The process of mapping meaning onto a quad in the form of a discursive concept is therefore something more than the sum of the lemmas in the quad. This process is thoroughly detailed in Fitzmaurice (forthcoming), and explicated in practice below.

In this qualitative examination, each process of analysis consists of hypothesis-building and iteratively affirming the legitimacy and generalisability of the hypothesis. To this end, we analysed examples until we were able to begin to reasonably generalise a hypothesis across the data. Two analysts independently analysed examples of quads, ranked according to the highest number of instances of its component parts within the 101-token span. Analysts each assessed the examples following the stepwise procedure outlined above, and assigned to each example preliminary descriptors of the characteristics of a discursive concept, with particular attention to religious or secular elements. We were then able to review our preliminary analysis and begin to generalise, aggregate, and identify trends in the data, as discussed below. We assessed our hypothesis to be reasonably generalizable after examining 70 examples and comparing and discussing analyses. Proposed next steps and future work are explored in the discussion section following the analysis.

#### 4. Analysis

We begin with an extensive analysis of an early example of the quad, from 1548. The stretch of text in Example (1), below, contains the four lemmas in close proximity, in a manner that we have observed to be highly frequent in EEBO-TCP, paraphrased many times across the decades. As far as the structure of the text is concerned, the lemmas occur in a quotation presented as a commandment (italics have been added to emphasise quad components here and in subsequent examples):

- (1) that the Gospel maye be sincerely taght preched for yt was the commandment which Christ gaue vnto his Apostles before he ascended vnto his father saying al *power* is now geuen me in *heuen* and in *erth* go therefore saith he into al partes of the *worlde* and preache the gospel (Kethe, 1548)

The first step in our analysis is semasiological inspection of each element to ascertain the most likely dictionary sense or sub-sense depending upon the immediate co-text. *Power* has several senses that might be relevant for this period. The primary one is the quality or property of mental or physical strength, vigour, might, energy or ability to act or affect something (OED, s.v. *power* (I.1.a)), which, despite this range of senses, remains underspecified without additional context. The text indicates that the quality of *power* is attributed to Christ by God (*his father*); accordingly, *power* might indicate an ability to cause an effect, illustrated in the OED as including the power of the lungs to move air, but also as in the power of Christ's blood to remove sin (OED, s.v. *power* (I.1.a)), which is at least obliquely evoked here. To the extent that Christ's abilities are aligned with the essential character of cause and effect in the world, the basic sense of cause and effect is at least somewhat present. That this range of sub-senses overlaps is already apparent. Moreover, *power* could indicate political or other authority, or rule over others, which might be given or conferred by some source (OED, s.v. *power* (I.1.b)), which is evoked here as well, insofar as Christ's authority is conferred by God. Of course, the nature of *power* in this narrative opens up centuries of discourse around power as a divine force, or a natural force, as political, or as personal, among many other underspecified elements.

*Heaven* could indicate the sky, whether near (containing clouds, rain, birds; OED, s.v. *heaven* (2)) or far (containing the sun, moon, and stars; OED, s.v. *heaven* (1)), with a sense of vastness (OED, s.v. *heaven* (1.c)), which is implied here. However, the reader is drawn instead to identify *heaven* as the abode of God and the angels, and the souls of the deceased; or, in parallel, a state or condition in which souls are with God in everlasting life (OED, s.v. *heaven* (5)). *Earth* could indicate the land and sea, in contrast to the sky (OED, s.v. *earth* (II.8)); or the

dwelling place of humankind, in contrast to the abode of God and the angels (and the souls of the dead; OED, s.v. *earth* (II.9.a)); or indeed the matter from which the human body (as opposed to the soul) is made (OED, s.v. *earth* (IV.14.a)). The juxtaposition of *heaven* and *earth* does nothing to clarify these ambiguities and underspecifications. The vagueness associated with the juxtaposition of the two terms is exacerbated by the fact that they occur so frequently in conjunction as to be a fixed phrase that defies easy, transparent interrogation. Discourses around *heaven* and *earth* (which abound) are therefore rich ground for manipulating the many vague and polysemous intersections of these spiritual and material senses. The trio *power-heaven-earth* is frequently instantiated in the phrase, *power over heaven and earth*, in which the relationship of *power* to the antonymic pair *heaven-earth* is one of dominion, the dominion of God (and Christ).

The fourth member of the quad, *world*, is both highly polysemous and vague. At the vaguest primary level, *world* denotes “the state or realm of human existence on earth” (OED s.v. *world* (I.1.a)), which on closer inspection is akin to the notion of the human lifetime. *World* could thus indicate the realm of humankind, sometimes conventionalised to refer to material pursuits or worldly affairs, with the post-Edenic implication of sinfulness; or indeed humankind or society itself. In analysing Example (1), particularly its use of *world*, it is necessary to review the prefatory material to the quotation in which the quad appears to begin to resolve the reference. This stretch of text has a bookend structure; it begins with the assertion that the Gospel can be properly taught on Christ’s commandment to his apostles on the strength of the authority conferred by his father; it ends with Christ’s imperative to his apostles to go to all the parts of the world to carry out this commandment: to teach the Gospel. Christ’s apostles are human inhabitants of the physical world, and so *world* appears to indicate that part of the earth that is the abode of humans (OED, s.v. *world* (II.6.a)). If the apostles are conferred Christ’s authority in the human realm to teach the Gospel, thus embodying God’s dominion over the world, it is reasonable to conclude that the fourth member of the quad contributes to the discursive concept about the power of God over everything everywhere.

The narrative of Example (1) is itself presented as true, with the authority of gospel; the powers of Christ in the narrative are true; and the commandment to preach this truth is a virtuous one to be upheld. It is interesting, perhaps, that the powers of Christ are not explicated here, but they would have been familiar to contemporary Christian readers.

The text in Example (1) is an early instantiation of a common trend for this quad in our data, with a standardised and relatively fixed interpretation (and thus also a culturally and socially salient interpretation for early modern writers and readers). It is formulaic, as scriptural language so often is when quoting or para-

phrasing biblical text. In brief, it captures the dictum that Christ has been given power over heaven and Earth to direct his gospel to be disseminated across the world, and is based on the Gospel of St. Matthew (18:18–20). References to this gospel, both oblique and direct, and paraphrases of the verse, occur repeatedly in the data, and might readily be seen as abstract or pure presentation of Gospel. Indeed, we would posit that writers more readily recycle content words such as these when they are salient due to their co-occurrence in biblical quotes or scriptural paraphrase and, as our analysis below illustrates, this ready recycling extends to non-religious domains (though it might be said that all domains are religious in Early Modern English). The vagueness in these content words renders them particularly useful for such re-use, as they remain salient but semantically malleable; their repetition in biblical quotation seems to do nothing to reduce that vagueness. The reading of this quad in this particular text, namely, <divine power over heaven and earth> invites negotiation (and manipulation, and evaluation) of the nature and extent of authority (and even ability) in worldly domains and beyond, as we add further context.

Additional context complicates the simple reading elaborated above. While the discourse that contains the quad is to be evaluated as true and right, the writer's aim goes further, as shown in the lines that follow Example (1):

- (2) ...but we fynde not in all the whole scripturs that euer he commanded any of them to say masse or to set vp stockes stones to make the peple beleue in the fylthy tradicions of mēs inuenciōs, as water, salt, bedes bowes, belles, with such other bagag and therefore ought all true Christē men without ceasing to praye vnto the lorde that he wold vouchsafe of his greate mercy to helpe and strenghten our noble kynge with all his godly counselors in takinge away of all suche popisshe constitucions which hath bene the destruction of many soules and to the great hindaunce of his most blessed word for the encrease nowe wherof let vs all with one mynde praye vnto god ye father of our lord Iesus christ. whose pease I hartely wysh vnto al thē yt vnfaynedly loue his testimonies. (Kethe, 1548)

The addition of further text radically destabilises the interpretation arrived at through the use of local context to produce an interpretation of the quad as a (neutral, agnostic) concept denoting the omnipotence of God. Example (1) contains implicit Biblical quotation in order to invoke the gospel according to St Matthew. We thus infer that Christ's injunction to his apostles to spread his word is true and righteous. Additional context allows consideration of the rhetorical structure of the larger text (Example (2)) in which the writer deploys the rhetorical ploy of partial quotation in order to attack a position or state of affairs he marks as contravening the true and righteous. In Example (2) the writer affirms

the *power* in Example (1) as conferred by God, in order to argue against the legitimacy of the authority, practices, and traditions of the Catholic church. This challenge is hinted at in the choice of the phrase *say masse* to contrast with the teaching of the *Gospel* in the preceding passage; this choice is strengthened in the emotive phrase *fylthy tradicions* attributed to mens' inventions rather than to God. This further informs our reading of *world*, which – encompassing *tradicions*, beads, bows, and bells – can now be interpreted with some element of its corrupt and sinful sense.

The juxtaposition of these discourses of righteousness (of the Gospel) and worldly wickedness (of the Catholic church) does not change the reading of the discursive concept that maps onto the quad. However, it does indicate the extent to which a ubiquitous and highly salient Christian notion can be appropriated for particular purposes, and re-cycled to create or contribute to timely contemporary discourses because of the semantic, pragmatic, and discursive malleability allowed by the underspecificity in these terms. The text illustrated in Examples (1) and (2) comes from a short pamphlet entitled *A ballet declaringe the fal of the whore of Babylone intytuled Tye thy mare from boye w[ith] other and there unto annexed a prologe to the readers* (Kethe, 1548). The use of the *whore of Babylon* image to refer to the Catholic Church indicates the violence of the author's antipathy to that institution and indicates the extent to which the author's invocation of St Matthew's Gospel is designed to support a violently ideological and partisan attack. Accordingly, although the stance of the text is clearly protestant and thus ostensibly religious, the force of the author's exegesis is sectarian and thus political. A very old discursive concept has been harnessed to generate a very new and timely discourse.

Example (3) illustrates a different instantiation of the quad; here, *power* appears three times compared with the single occurrence of each of the other members of the quad. This fact insists upon the fresh inspection of the meanings of each occurrence and its relation to the interpretation of the quad as encountered in Example (1). The instances of *power* can be construed as indicating ability and authority, given from God to Christ and from Christ to the apostles, in *Heaven* and *Earth* (here underspecified), but, in a significant addition, *power*, in its lexical semantics, indicates both ability and authority (again underspecified) of every man to interpret scripture. In its first occurrence in Example (3), *power* is an ability natural to the human intelligence – but also (implicitly) granted to men by God, with effects on (or dominion over) the worldly and the spiritual.

- (3) If the *power* of Interpreting of Scripture be in every mans brain, what need have we of a Church or Church-men? To what purpose then did our Saviour, after he had given his Apostles *Power* to Bind and Loose in *Heaven* and *Earth*, add to it, that he would be with them even to the End of the *World*? These

words were not spoken Parabolically, or by way of Figure. Christ was then ascending into his Glory, and left his *power* with his Church even to the End of the *World*.  
 (Charles II, 1685)

The quad appears as a straightforward reference to the discursive concept <the power of God over heaven and earth>, as encountered in Example (1). In its second occurrence in Example (3), *power* is – as in the scripture of Example (1) – granted by God to Christ, again with effects on (or dominion over) the worldly and/or spiritual domains. However, in this text, the writer refers to Christ’s delegation to his apostles of his *power* to “bind and loose in *Heaven* and *Earth*”. To investigate what this power consists of, we need to return to the gospel for context, back to Matthew. This additional context helps to solve the vagueness and under-specification of the textual relation between *power*, *heaven* and *earth*, indicating the radical truncation of the discursive frame for the trio. It can be elaborated with reference to Matthew 16: 19: “And I will give unto thee the keys of the kingdom of heaven: and whatsoever thou shalt bind on earth shall be bound in heaven: and whatsoever thou shalt loose on earth shall be loosed in heaven” (The Bible, KJ version 1611). The writer’s comment on Christ’s promise to remain with his Apostles *even to the End of the World* that these words were not spoken using parable or metaphor, is designed to assert the presence of Christ’s power, authority, in the church on earth until the end of the world. Additionally, the fourth member of the quad, *world*, is not linked to the other members in the way it is in Example (1) (and in Matthew 28). Here, the two instances of *world* refer to the state or realm of human existence on earth (OED s.v. *world*, I.1.a), with no certain implication of sin or corruption.

Of course, the first occurrence of *power* in Example (3) is attributed to humans rather than God – the writer employs the word *power* to refer to the ability of every man to interpret scripture, but argues that that ability, on its own, is insufficient and likely to lead to error. Like Example (1), this Example is polemical; unlike Example (1), it is not the Catholic church that the writer opposes, but the *protestants* opposed to the authority of the Church of England. Thus, this example emphasises the conferral of power not just by God to Christ, but by Christ to the apostles, and the apostles to the church (i.e. the Church of England). The example is drawn from the letters of King Charles II, re-printed multiple times in EEBO-TCP; the author juxtaposes, then, a falsely claimed authority of every man to interpret scripture, and a true authority held by the Church of England. This additional historical (and political) context allows further clarification and specification: Christ’s church on earth was Peter; in the Stuarts’ time, Christ’s true church on earth is the Church of England, that is, the Church of Charles II and his brother, James II. It is the reference to the gospel of Matthew,

as signaled by the quad, that forms the salient background or frame, or reference, for this argument.

Example (3) exemplifies how underspecification in language, and vagueness in discourse, can be a catalyst, or an invitation, or at least an opportunity for a writer to re-interpret, redefine, and re-create key social and cultural concepts, in context. The discursive meanings of the members of the quad are remarkably complex, and the single lemma *power* is employed three times in quick succession with inconsistent senses or sub-senses. However, this variation is unlikely to provoke serious problems or miscommunication for readers then or now, even if it is much murkier below the surface.

Many instantiations of the quad appear in specific descriptions of God's omnipotence in terms of its strength and its vastness. The following examples also illustrate the complexity of *power* and show the critical role of context in clarifying its meaning and relationship with the other members of the quad:

- (4) That the Whole *World* does Vnseen *Powers* obey? Those Vnseen *Powers* that  
act what e're they please, And Hidden, rule *Heav'n*, *Earth*, and Air, and Seas;  
Which trembling Human Race in Prayer adores (Barker, 1698)

At first glance, the meaning of *power* in Example (4) appears to be in sharp contrast to *power* in Example (3), as human ability and authority: *power* here is *unseen* and *hidden* (forces), inducing human *trembling* and adoration. We might initially be encouraged to read this text as heralding the immensity of the divine forces that rule the entire human and physical world. However, the co-text challenges this reading; the line preceding the text is: "O my astonish'd Muse what can we say But that we know upon mature Survey" and the line following it is: "And Nassau only imitates in Secret Powers". This suggests now that the miraculous immensity of divine power provides the backdrop for the ecstatic celebration of the achievements of *Nassau*. The bibliographic and historical contexts of the text itself shows that the excerpt is from a poem to commemorate the Treaty of Ryswick, which ended the Nine Years' War between France and a European coalition including England. The peace was negotiated in 1697, by William, Prince of Orange, also King William III of England. *Nassau* is the pseudonym given to the coalition's hero and the poet invokes the omnipotence of God in order to amplify *Nassau's* awesome mastery over the *Numerous Foes*. So the concept of divine omnipotence, measured as dominion over heaven, earth and the human world, is invoked to glorify the achievements of men here – again a sharp contrast with previous examples, and a different application of the quad in a different discourse.

Likewise, Example (5) indicates the immensity of scale insofar as earth is *his footstool*.

- (5) He who rideth on the skies, and makes the *earth* his footstool; that is Alpha and Omega, the beginning and the ending. It was this good God, that first fram'd this Universe; It was the Lords work, and the Lords will that created this *world*: It was the glorious almighty *power* of *Heaven* that first founded this fabrick (N.A., 1658)

In both Examples (4) and (5), *world* is no longer the geographical expanse for evangelism, as in Example (1), nor a corrupt and sinful realm, as in Example (2), but rather the whole of existence that “obeys” and was created by God’s vast *power*. These instantiations of the quad (also quite common) do not cite the scriptural backdrop of Matthew, but can be seen as intensely spiritual or religious, indicating the miraculous rather than the secular.

The examples above particularly emphasise the complex discursive relationships between the quad members, although the pair *heaven* and *earth* so far appears to be very stable (thus compensating for their underspecification). *Power* is a dominant component; as well as occurring frequently around the quad, its meaning is constantly open to interpretation, requiring careful scrutiny of its relation to the other quad members. The consequence is that it varies subtly between power as divine omnipotence; divine authority; and human authority, capacity and ability; and the relationship between the divine and human senses, insofar as human authority or ability may or may not be granted by God. In line with this malleability, the fourth member of the quad, *world*, refers to the entire domain of God’s omnipotence, to human existence on earth, or indeed the collective human habitation on earth, with or without an element of sin and corruption.

Let us explore further the nature of the apparent underspecification of *heaven* and *earth* in Examples (6) and (7):

- (6) not onely this Nations wisdom in their very supream Assembly but the Ruling *powers* of the whole *world* shall yield subjection and think it their glory to become subservient to the bringing forth of the new *Heavens* and that new *Earth* wherein dwells righteousness and truth (Vane, 1656)

The *new heavens* and *earth* promised signal a departure from the highly stable pair encountered in the instantiations of the quad thus far. The suggestion that heaven and earth can be renewed or replaced raises questions about the truth and permanence of what we might have understood to be unshakeable. The inspection of the sense of *powers* helps to clarify this revelation. *Powers* here denotes political powers, in the form of states (OED, s.v. *power* (II.8.c.)), a discretely polysemous meaning from those described above, even if it overlaps with questions of (legitimate) conferral of authority, and strength and ability to cause effects in the world. As *power* is re-located in the human realm, so the pair *heaven* and *earth* is destabilised.



Again, to resolve the mystery of the *new heavens and earth*, it is crucial to seek context to enrich the analysis of the relationships between the quad members. This text is taken from the final testimony of Sir Henry Vane, prosecuted and imprisoned by the Protector, Oliver Cromwell, for sedition in 1655, “tending to the disturbance of the present Government and the peace of the Commonwealth” (Vane, 1656). We make sense of the assertion that there will be brought forth new heavens and earth in the context of the political battle for legitimate definition and authoritative possession of the power bestowed on earthly representatives by God on the one hand and for the true interpretation of the power of God on the other.

In Example (7), we again see the *new heaven and earth*, but juxtaposed with the *whole Heavens and Earth*:

- (7) The Indignation of the Lord is upon all Nations, and his Fury upon all their Armies; he will greatly distress the Inhabitants of the *World*; he will now plead the Cause of his Covenant of Light against all those that withstand his Glorious Appearance, and Work of Restoration, in this the Day of his Mighty *Power*, wherein he is giving the Kingdom of his Everlasting Peace to *Israel's* Seed under the whole *Heavens*, who fear his Great Name, and tremble at his Almighty Word of Life, which shakes not only *Earth* but *Heaven* also: And now is the Lord God Almighty bringing forth his Glory to light, and creating the *New Heavens* and *new Earth*, wherein dwells Righteousness according to his Promise, *Isa 65.17, &c.* (Taylor, 1667)

In this text, the relationships between the members of the quad seem to have shifted. However, at the heart of the text is the discursive echo of the Matthew gospel; *power* is the might and authority that is unmistakably God's, which he has given to Jesus Christ and through him to his church on earth. However, there is a different relationship between the lemma *world* and the pair *heaven-earth*; God's *indignation* and *fury* will *distress the Inhabitants of the World* in the present dispensation under heaven and earth. Here the world is that earthly domain of corrupt and sinful nations and their armies which God's power must cleanse and deliver. The writer then explicitly invokes Isaiah's prophesy of Christ's coming as God's *bringing forth his Glory to light*, and the creation of new Heavens and new Earth. Again, information afforded by bibliographic and historical context helps to clarify the way in which the old discursive concept is repurposed for active religious protest. The author of this tract was Thomas Taylor, a prominent Quaker minister and writer who was continually imprisoned for refusing to swear the oaths of allegiance in the Stuart monarchy (Fell-Smith, 2004).

The quad is also unsurprisingly employed in discourses that emphasise political authority with only peripheral reference to the religious.

- (8) The next thing to be considered is what *power* either God or man hath given to the Magistrate in things spiritual or meerly relateing to conscience for this is most certain if neither have given him *power* he has none but what he assumes of himself which is a plain intrusion and so the exercise of it must needs be unlawful But before we proceed to determine the question we must first take notice that God has given all *power* in *Heaven* and in *Earth* to his Son the Person of Christ for none can pretend to any *power* whether in civils or spirituals which he derives not from him in civils we have already shewn that there are subordinate Magistrates set up in the *world* under Christ the great Potentate and King of Kings to whom we are to be subject for conscience sake being an Ordinance of God Rom. 13.2.5 in spirituals is now the question what *power* Christ has derived to man or whether he has reserved the sole *power* to himself in things that meerly relate to conscience wherein we are to stand and fall to our own Master and not to be judged by others (N.A., 1659)

In Example (8), the writer cites and interprets scripture to describe and explain *power* of civil magistrates – but it is left explicitly underspecified whether authority is granted by God or by humanity. That is, the writer explicitly presents *power* as authority conferred by *either* God or man, and then cites the gospel of Matthew as the relevant frame, or background. Unlike Example (3), which explicitly argues that authority conferred by men to men, over the spiritual domain, is illegitimate and unjust, Example (8) allows complementarity, such that *power* might be granted by either, neither, or both.

Example (8) is drawn from a lengthy and complex argument about the nature and scope of civil political authority, in relation to religion and ‘conscience’ – i.e. a conscious sense of morality (OED, s.v. *conscience* (I)) or, as described in this same text, a mental understanding of God. The argument can be seen as philosophical and political, but also as practical and administrative. This discourse, then, contrasts with the more general citation of scripture in Example (1); like Example (3), the discourse in Example (8) uses scripture as the premise of a practical argument about the self and society, but extends further into the administrative and secular domains. Crucially, *world* here encompasses the administrative and secular human realm, and is seen as virtuous, right, and just, rather than sinful and corrupt. And it is clear that Examples (7) and (8) are in stark contrast to those texts that illustrate overwhelming divine might in the form of miracles.

In Example (9), we again see a political argument and we observe the use of a fifth lemma, *law*, alongside *world-heaven-earth-power*, as an expression of just authority in the worldly domain (similarly, Example (8) above included *unlawful*).

- (9) The True Light which Enlighteneth Every Man that cometh into the *World* To HIM as Made Man all *Power* in *Heaven* and *Earth* is Committed for the Exe-

cution of the Office of the MEDIATOR between GOD and Man Most certainly His *Power* is Infinitely Superior to that of King Lords and Commons But it is an Act of Disobedience to Him to Refuse our Active Obedience to any Law of the Land which we Know or may easily Know that we may Observe without any Repugnancie to the Will of GOD Revealed in the Holy Scriptures  
(Elys, 1690)

*Law* in Early Modern English could indicate God's commandments (OED, s.v. *law*, (I.9.)); or any rule imposed by authority (OED, s.v. *law* (I.)); or an accepted norm not imposed by authority (OED, s.v. *law*, (II.14.1)); or, from the 17th century, some convergence of all three; so it is unsurprising that it would be employed repeatedly within this quad's cotext. By the late 17th century, *law* could refer to facts about the universe, and cause and effect, often in *laws of nature*, *laws of physics*, and so on (OED, s.v. *law* (III.17.a.)). As we have shown, *power* can be legitimate or not, but in repeated examples that we have observed alongside this quad, *law* is specifically employed to represent legitimate, just, and right rules in the political, secular domains. This would seem to be in accordance with the contemporaneous emergence of *laws of nature* as correct truths, viewed from a relatively secular perspective.

As the examples here demonstrate, the underspecification in the lexical items in the quad facilitates a range of negotiated and manipulated discourses. In particular, we have identified two crucial, prominent discursive concepts. The first might be expressed as <God is omnipotent>; a prominent discursive concept built upon the Gospel of Matthew, and the relatively fixed narrative of God giving Christ power over heaven and earth, across the world. The second discursive concept appears to be marked by a discursive tension between the trio *power-heaven-earth* and the fourth member of the quad, *world*. In a number of examples, two kinds of power are invoked; the first is God's omnipotence. The second is man's translation or management of God's power in the world of men over men, through the laws of institutions such as monarchies or churches. This is a more complex discursive relationship as it pivots on the juxtaposition of God's power with the authority assumed by men over men in the world, an authority which is contested among men. The contestation is played out in sectarian terms, as witnessed in the text in Example (3). These secular, administrative and political aims constitute newly emerging practical negotiations for which the frame of God's omnipotence is foundational. This second discursive concept may be a creative negotiation or manipulation of the first; the second harnesses the power of the first even as it pushes against it. The aims of the second are not strictly spiritual or exegetical; they connect to scripture, but none of the examples is purely a sermon, or scripture, or exegesis – and they are published as pamphlets, commentaries, or testimonies that use scripture as a premise for advancing an argument

about authority. The primary discursive concept – which is uncontested – may be <God is omnipotent>, but it functions in discourses around the translation of God's power in the (often corrupt, sinful) world. And we observe echoes of this from 1548 in anti-Catholic texts; in the restoration; and in the intensely sectarian content in the Quaker debate, which extends to Presbyterians and others as well. We might interpret the second discursive concept as a move towards more secular instantiations of this quad, out of the salient building blocks of *earth-heaven-power-world* that were already available – and malleable – via the first discursive concept, and the emerging reliance on *law* seems to reflect that.

We hypothesise that the two discursive concepts might represent change in the quad over time; we arrived at this hypothesis iteratively, through close semantic and pragmatic analysis of nearly 100 instantiations of the quad. To affirm change over time will require quantitative analysis of thousands of examples, using methods similar to those here; those next steps are beyond the scope of the present work.

## 5. Conclusion

We have illustrated how writers employ the quad *earth-heaven-power-world* in a range of discourses, to negotiate sometimes charged social and cultural issues and meanings. These discourses include established religious discourses, as well as newer, practical, secular discourses that nonetheless show a strong tendency to cite those older, established religious discourses. The older discourses constitute a sort of historical backdrop against which other, newer discourses can be measured – and against which relevance or salience of interpretation might be assessed. Divergence from this historical standard might then loosely be seen as an indicator (or reflection) of discursive change. When an established backdrop like the gospel of St Matthew is appropriated to justify contemporary political arguments, we observe a synthesis of ideas that results in the emergence of a new discourse.

We have also demonstrated that underspecification in these terms is a critical ingredient in the emergence of new discourses. That underspecification is a source of semantic change is already well understood, if under-emphasised. Semanticists since the 19th century have understood that semantic change occurs when meaning that is not expressed semantically (in the word) is instead specified pragmatically (in context) and that such pragmatic specification can be conventionalised into semantic meaning. For example, in the philological tradition, Schleiermacher posited such a mechanism of change in the first half of the 19th century, which Paul took up in the second half of the 19th century (see Nerlich,

1992: 31–32; Geeraerts, 2010: 14–16). More recently, in the cognitive tradition, it has been shown that when a lexeme undergoes semantic change, it tends to first convey different meanings in different contexts via specific contextual enrichment (Traugott & Dasher, 2002). What is generally less explicit is that any underspecified element of lexical semantic meaning can be specified pragmatically and that, in turn, it is vagueness – as underspecification – that is very often the site or source of semantic change (see Mehl, 2012). We have shown here that extremely common discourses, and charged debates, can be built on constellations of non-adjacent, co-occurring semantically underspecified lexemes, in varied relations with each other and with their contexts. As social and cultural contexts change, and as charged debates erupt and subside, the semantic and pragmatic relationships between underspecified lexemes can be harnessed or appropriated to give rise to new and changing discourses. If vagueness is the site and source of semantic change, then it is accumulated vagueness across constellations of lexemes such as the quad examined here that is a rich site and source of discursive change.

Moreover, we have illustrated just how valuable quad data is as a unique way in to observing discursive variation and change, by facilitating semantic and pragmatic analysis of discourse in a unique way. We would contrast this with more established methods. For example, the analysis of topical or thematic corpora begins with corpus construction: Baker et al. (2013) analysed perceptions of Islam and Muslims by identifying a word list (largely synonyms) and automatically extracting a large number of news articles containing those words. Conceptual historians (see Koselleck, 2004) often begin with a keyword and manually branch out to related words. Our method begins not with an onomasiological concept like Baker et al.'s (2013), nor with a semasiologically based conceptual structure like Koselleck's (2004), but rather with a set of strong discursive relationships between lexical items. Other methods that rely on sets of lexical items are in fact far more likely to systematically disregard underspecification or vagueness: for example, topic modelling and distributional semantic techniques tend to algorithmically sort out those words that are used in a wide range of contexts, to focus on a short list of words with narrower, more precise meanings. Our method, in contrast, employs computation to embrace underspecification, and applies semantic and pragmatic analysis to interpret accumulated underspecification as discursive change. If we were only looking at power, and investigated *power* as a search term, we would not be able to observe the complexities of these lexical relationships, and we might never ascertain the importance of the gospel of Matthew as a discursive frame, which echoes through early modern debate.

## Funding

Open Access publication of this article was funded through a Transformative Agreement with University of Sheffield.

## References

- Baker, P., Gabrielatos, C., & McEnery, T. (2013). Sketching Muslims: A corpus driven analysis of representations around the word 'Muslim' in the British press 1998–2009. *Applied Linguistics*, 34(3), 255–278. <https://doi.org/10.1093/applin/ams048>
- Barker, T. (1698). *Nassau: A poem occasioned by the peace*. Text Creation Partnership.
- Burns, P. (2013). *MorphAdorner* (Version 2) [Computer Software]. Northwestern University. <https://morphadorner.northwestern.edu>
- Charles II (1685). *Copies of two papers written by the late King Charles II of blessed memory*. Text Creation Partnership.
- Cruse, A. (2011). *Meaning in Language: An Introduction to Semantics and Pragmatics*. Oxford University Press.
- Elys, E. (1690). *Animadversions upon some passages in a paper, entituled, Some reasons humbly offered why the Quakers should be examined*. Text Creation Partnership.
- Fano, R. M. (1960). *Transmission of Information: A Statistical Theory of Communications*. MIT Press.
- Fell-Smith, C. (2004). Thomas Taylor 1617/1618–1682. *Dictionary of National Biography*. Oxford University Press.
- Fitzmaurice, S. (forthcoming). From constellations to discursive concepts; or: The historical pragmatic construction of meaning in Early Modern English. *Transactions of the Philological Society*.
- Fitzmaurice, S., Robinson, J. A., Alexander, M., Hine, I. C., Mehl, S., & Dallachy, F. (2017). Linguistic DNA: Investigating conceptual change in Early Modern English discourse. *Studia Neophilologica*, 89, 21–38. <https://doi.org/10.1080/00393274.2017.1333891>
- Foucault, M. (2003). *Society Must be Defended: Lectures at the Collège de France, 1975–1976*. Translated by David Macey. Picador.
- Geeraerts, D. (2006). Vagueness's puzzles, polysemy's vagaries. In D. Geeraerts (Ed.), *Words and Other Wonders* (pp. 99–148). Mouton de Gruyter. (Reprinted from *Cognitive Linguistics*, 4–3, 223–272, 1993) <https://doi.org/10.1515/9783110219128.2.99>
- Geeraerts, D. (2010). *Theories of Lexical Semantics*. Oxford University Press.
- Hoey, M. (2004). Lexical priming and the properties of text. In A. Partington, J. Morley, & L. Haarman (Eds.), *Corpora and Discourse* (pp. 385–412). Peter Lang.
- Hoey, M. (2005). *Lexical Priming: A New Theory of Words and Language*. Routledge.
- Kethe, W. (1548). *A ballet declaringe the fal of the whore of babylone intytuled Tye thy mare tom boye w[ith] other and there vnto anexid a prologe to the reders*. Text Creation Partnership.
- Koselleck, R. (2004). *Futures Past: On the Semantics of Historical Time*. Columbia University Press.
- McEnery, M., & Baker, H. (2017). The poor in seventeenth-century England: A corpus based analysis. *Token: A Journal of English Linguistics*, 6, 51–83.

- Mehl, S. (forthcoming). Discursive quads: New kinds of lexical co-occurrence data with linguistic concept modelling. *Transactions of the Philological Society*.
- Mehl, S. (2019). Measuring lexical co-occurrence statistics against a part-of-speech baseline. In H. Parviainen, M. Kaunisto, & P. Pahta (Eds.), *Corpus Approaches into World Englishes and Language Contrasts*. VARIENG. <http://www.helsinki.fi/varieng/series/volumes/20/mehl/>
- Mehl, S. (2012). Thinking linguistically about keywords: Polysemy, semantic change and divergent identities. *The Keywords Project*. [https://keywords.pitt.edu/pdfs/thinking\\_linguistically\\_about\\_keywords.pdf](https://keywords.pitt.edu/pdfs/thinking_linguistically_about_keywords.pdf)
- N. A. (1658). *Londons wonder being a most true and positive relation of the taking and killing of a great whale neer to Greenwich; the said whale being fifty eight foot in length, twelve foot high, fourteen foot broad, and two foot between the eyes. At whose death was used harping-irons, spits, swords, guns, bills, axes, and hatchets, and all kind of sharp instruments to kill her: and at last two anchors being struck fast into her body, she could not remoove them, but the blood gusht out of her body, as the water does out of a pump. The report of which whale hath caused many hundred of people both by land and water to go and see her; the said whale being slaine hard by Greenwich upon the third day of Iune this present yere 1658. which is largely exprest in this following discourse*. Text Creation Partnership.
- N. A. (1659). *The true magistrate, or The magistrates duty, and power in matters of religion, or in things meerly relating to conscience. By a well wisher to a publick good*. Text Creation Partnership.
- Nerlich, B. (1992). *Semantic Theories in Europe 1930–1930*. John Benjamins. <https://doi.org/10.1075/sihols.59>
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford University Press.
- Tadmor, N. (1996). The concept of the household-family in eighteenth-century England. *Past and Present*, 151(1), 111–140. <https://doi.org/10.1093/past/151.1.111>
- Taylor, T. (1667). *A vvarning to the nations to lay aside all prejudice and enmity the ground of strife and wars and to come and embrace the light, Christ Jesus*. Text Creation partnership.
- Traugott, E. C., & Dasher, R. (2002). *Regularity in Semantic Change*. Cambridge University Press.
- Vane, H. (1656). *The PROCEEDS of the PROTECTOR (so called) and his COUNCILL against Sir HENRY VANE, Knight. A late Eminent MEMBER of the COUNCILL of STATE of the Common-wealth of ENGLAND: As touching his Imprisonment in the Isle of WIGHT. Together with his Testimony delivered in writing to the said Protector, by way of a peculiar addresse to himself, and faithfull advice of an old Friend and Acquaintance. Faithfully Transcribed and Attested out of the Original Copy's, And published by a reall Well-wisher to SIONS Prosperity and ENGLANDS Liberty*. Text Creation Partnership.

## Address for correspondence

Susan Fitzmaurice  
Faculty of Arts and Humanities  
University of Sheffield  
34 Gell Street  
Sheffield S3 7QY  
United Kingdom  
s.fitzmaurice@sheffield.ac.uk  
 <https://orcid.org/0000-0002-8804-1987>

## Co-author information

Seth Mehl  
Digital Humanities Institute  
University of Sheffield  
s.mehl@sheffield.ac.uk  
 <https://orcid.org/0000-0003-3036-8132>

## Publication history

Date received: 6 October 2020  
Date accepted: 14 January 2022  
Published online: 6 September 2022