

Looking for concepts in Early Modern English

Hypothesis building and the uses of encyclopaedic knowledge and pragmatic work

Susan Fitzmaurice
University of Sheffield

The idea that conceptual meaning in discourse could be identified in constellations of lexical co-occurrences in a particular “universe” of discourse was key in guiding the computational historical semantic–pragmatic work conducted in the Linguistic DNA project. The project mapped prominent lexical co-occurrences across the two hundred years of publications in Early English Books Online (EEBO-TCP; Text Creation Partnership edition), yielding concept models – constellations of non-adjacent lemmas that consistently co-occur across spans of up to 100 tokens. The goal was to map meaning onto concept models as “discursive concepts”, using encyclopaedic knowledge, pragmatic analysis and context.

The first question concerns the effectiveness of making early hypotheses about the discursive meaning of concept models based on the inferred connections between the lemmas in a quad constellation. The second question is whether the meaning of frequent, apparently stable concept models changes upon their closer scrutiny in the discourses they lead us into. A reader familiar with the particular universe of discourse in which these quads occur, and with the social, historical, literary and philosophical traditions, and the context that they occupy, might be effectively primed by their encyclopaedic knowledge to hypothesise this discursive meaning. This paper demonstrates the efficacy of hypothesis building using encyclopaedic knowledge and pragmatic analysis to interpret optimally relevant concept models.

Keywords: concept models, discursive meaning, Early Modern English, encyclopaedic knowledge, hypothesis building, pragmatics

1. Introduction

A major challenge posed by the big text data yielded by computational methods of data extraction and analysis is to bring its vastness back to human proportions. The Linguistic DNA project mapped prominent lexical co-occurrences across the two hundred years of publications in Early English Books Online (EEBO-TCP; Text Creation Partnership edition).¹ The output of the Linguistic DNA processor is billions of concept models – constellations of pairs, trios or quads of non-adjacent lemmas – that consistently co-occur across spans of up to 100 tokens. In simple terms, the concept modeller embodies a process of distant reading, to yield as outputs, vast numbers of constellations of lemmas, which we labelled concept models. These concept models provide a distinct route to identifying the kinds of discourses that make up the universe of early modern texts. The central concern of this paper is to examine how concept models signal particular discourses and then what rhetorical roles those concept models perform within those discourses.² The major issue is how readily and with what aids the reader can discern the relationships between lemmas in the constellations to hypothesise a meaning for the concept model itself.

The project has focussed on the analysis thus far of constellations that occur extremely frequently in EEBO-TCP in order to ascertain the extent to which they turn out to be the most prevalent indications of this universe of discourse. Fitzmaurice and Mehl (2022) and Fitzmaurice (forthcoming) explore how different highly frequently occurring quads acquire discursive meaning in their particular textual co-texts, applying a series of semantic–pragmatic routines to better understand the rhetorical work performed by the quads in the discourse. Fitzmaurice (forthcoming) argues that the meaning constructed from the quads in the documents in which they predominate is discursive rather than lexical; that is, although the discursive meaning is demonstrated to be highly stable across the

1. “Linguistic DNA: Modelling Concepts and Semantic Change in English 1500–1800” (2015–2018) is a collaborative research project funded by the UK Arts and Humanities Research Council (AHRC AH/M00614X/1). The Principal Investigator was Susan Fitzmaurice (Sheffield), with Co-Investigators Michael Pidd (The DHU, Sheffield), Justyna Robinson (Sussex) and Marc Alexander (Glasgow), assisted by three research associates: Fraser Dallachy (Glasgow), Iona Hine and Seth Mehl (Sheffield). Technical development was carried out by Matthew Groves and Katherine Rogers (Sheffield), and Brian Aitken (Glasgow). Further details are available online at: <http://linguisticdna.org/>.

2. The notion of discourse here refers to that mode of organising 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 (OED s.v. *discourse* 7).

genres of discourse in which they appear, that meaning evades simple or straightforward lexical expression. Fitzmaurice and Mehl (2022) show how the same concept models may perform different rhetorical roles depending on their textual context. For example, although the discursive meaning might appear stable over time and across documents, it is recruited to do different kinds of work in particular discourses. We demonstrate how biblical quotation, which hosts a quad, is wielded in a sectarian paper-war between protestants and Catholics in mid-seventeenth century England.

In this paper, I examine several quads that can be interpreted as concept models very easily, with what at first appears to be no context at all. It seems to me that a reader who is familiar with the particular universe of discourse (e.g., EEBO-TCP) in which these quads occur so markedly frequently, and with the social, historical, literary and philosophical traditions, and context that they occupy, might be primed by their encyclopaedic knowledge to hypothesise this discursive meaning. Another approach to interpreting concept models is to consider how the reader's immersion in, and saturation by, their cultural and intellectual context might make particular candidate meanings more relevant than others might be. The method for testing whether early hypothesis building is sufficient to assign meaning to quads involves assessing how far the resulting concept model retains its meaning within its discourses. By examining the impact of additional information and context on the stability of the concept model, we can judge the robustness of the concept model as proxy or shorthand for the discursive concept. Put another way, close historical pragmatic analysis of how far the meaning of the concept model responds to successive layers of context provides a method for assessing the effectiveness of hypothesis building.

In Section 2, I present the Linguistic DNA method used to interpret the concept models that mark out the discourses encountered in EEBO-TCP. Because the detail is amply described and justified elsewhere (e.g., Mehl, 2019; Fitzmaurice and Mehl, 2022; Mehl, forthcoming), and because I wish to explore what we can do with the concept models, I will focus on the output generated by the concept modeller (namely, the quads of four strongly linked lemmas). In Section 3, I look at two sets of quads as different cases of hypothesis building in mapping meaning to quads as concept models. These cases indicate that the most relevant discursive meanings of concept models appear at different stages of the process.

2. The analysis of LDNA processor output as concept models

The project developed a novel set of computational methods for calculating lexical co-occurrence in large spans of text to identify prominent conceptual

content in Early Modern English texts. The project considered using distributional semantics and collocational analysis but these methods were limited in being able to generate analysis beyond co-occurrence pairs. We discarded topic modelling because we needed to be able to locate and see linguistic examples of the conceptual content in the texts analysed. The process we call “concept modelling” was developed to locate conceptual content in groups of strongly associated lemmas within a span of text, whether the group is a pair of non-adjacent lemmas, a trio of non-adjacent lemmas or a quad of non-adjacent lemmas.

The input data for the concept modeller consists of the Text Creation Partnership’s (2015) edition³ of Early English Books Online (EEBO-TCP), which was hand-keyed by professional transcribers. There are more than 60,000 texts in the corpus, comprising documents printed in English or published in the British Isles between 1473 and 1700. To prepare the data for concept modelling, the project adopted MorphAdorner (Burns, 2013), which was used to regularise the spelling, lemmatise the data and Part-of-Speech (POS) tag the entire corpus, with an iterative process of manual correction. In the course of the continual development and refinement of the process, the project experimented with using grammatical stopwords and excluding from analysis exceptionally frequent lexical items, such as *god*. As the project experimented with the process, the input data varied; we ranged from limiting the candidate node words and constellation members to nouns, to including nouns, adjectives, verbs and adverbs to examine the kinds of concept models generated.

The output of the computational process is thus a list of quads identified, ranked and sorted, using quantitative measures. The strength of the co-occurrence relationship among lemmas is calculated using a statistical measure called Mutual Information score (MI), adapted for use in the project (Mehl, 2019). Since the process is, computationally, very expensive, very high frequency thresholds were applied, yielding quads around very high frequency node words. The quads can be ranked by frequency or by MI score; quads with both high frequency scores and high MI scores are classified as “prominent” quads. The format and explanation of this output using the Concept Modelling Demonstrator can be inspected on the project website.⁴

Table 1 illustrates how the processor output appears in the Demonstrator. The node word in this list of quads is *day*. The quads share some of the same members, units of time: *day*, *hour* and *minute*. The appearance of additional means of measuring time, namely, *degree*, the specific event measured, *eclipse*, and *moon*,

3. For details of how the Linguistic DNA project used the Text Creation Partnership’s 2015 edition of Early English Books Online, please see Fitzmaurice et al. (2017: 25–27).

4. See: <https://www.linguisticdna.org/>.

the basis of the lunar calendar, suggests a strong association among the lemmas in these quads. Appeal to encyclopaedic knowledge alone informs an early hypothesis that these quads might be a concept model signifying the notion of the lunar calendar.

Table 1. Illustrative concept modeller output for quads with the node word *day*

Word A	Word B	Word C	Word D	Freq.	MI score
<i>day</i>	<i>hour</i>	<i>minute</i>	<i>moon</i>	27,672	11.4679
<i>day</i>	<i>eclipse</i>	<i>hour</i>	<i>minute</i>	17,598	14.3206
<i>day</i>	<i>eclipse</i>	<i>minute</i>	<i>moon</i>	12,931	12.8749
<i>day</i>	<i>eclipse</i>	<i>hour</i>	<i>moon</i>	18,191	12.7500
<i>day</i>	<i>degree</i>	<i>hour</i>	<i>minute</i>	23,734	12.6443
<i>day</i>	<i>degree</i>	<i>long</i>	<i>minute</i>	12,081	12.6215
<i>day</i>	<i>hour</i>	<i>long</i>	<i>minute</i>	15,987	12.6107

Forming an early hypothesis in this way depends upon the availability of meanings that constitute the connections between the lemmas in a quad. The inference of hypothetical relationships between lemmas based on encyclopaedic knowledge (and the principle of optimal relevance) allows us to group particular lemmas in these quads into particular semantic fields without invoking any specific information. For example, *day*, *hour* and *minute* are related units of time. The lemmas that have the capacity to be more referentially specific, *eclipse* and *moon*, tend to be linked in a common phrase such as *eclipse of the moon*. These two groups can be connected straightforwardly, because taken together they might be relevant to the temporal setting of moon eclipses. The remaining lemmas, *degree* and *long*, are highly underspecified if each is considered in isolation. However, they can acquire meaning by virtue of their association with the other lemmas in their respective quads.

A reader familiar with the particular universe of discourse in which these quads occur and with the social, historical, literary and philosophical traditions, and context that they occupy, might draw upon their encyclopaedic knowledge and thus effectively be primed to hypothesise the discursive meaning around the lunar calendar or the cycles of the moon. We can inspect the document titles in which the quads in Table 1 occur to test this hypothesis. The exercise confirms that the quads relate to the concept of the solar year and its measurement using

the lunar calendar.⁵ There is, therefore, a very short inferential distance between inspecting the quads yielded by the processor and assigning meaning to them as a concept model. The concept model inferred from the quads surrounding *day* seems to rely primarily upon activating encyclopaedic knowledge rather than an iterative process of semantic–pragmatic analysis.

Now the level of specificity provided by the long title of the document in which the quads occur may be sufficient for those readers curious to discover what kinds of discourses make up the EEBO-TCP corpus. However, should we wish to explore what role the hypothetical concept model might perform within the discourses identified through the metadata consisting of the document titles, we can explore the documents themselves to scrutinise the quad as it occurs in the text itself. Example (1) illustrates the occurrence of the *day–eclipse–minute–moon* quad in its original co-text in a document.

- (1) THough the spring be seasonable, yet diseases proceeding from abundance of blood and choller are rife. Authority somewhat molested by men pretending holiness at home, yet curteously treated from forreiners. The six and twentieth **day** there hapneth a great **Eclipse** of the **moon**, the beginning at thirty four **minutes** past six in the evening, she riseth eclipsed: the greatest obscuration will appear at eight; the end will be somewhat after nine: the whole duration will be almost three houres: almost nine digits or parts will be darkned.

The quad appears in the forecast of a spring eclipse of the moon and the precise time and duration of the eclipse. Inspection of the quad within its discourse thus confirms the meaning of the concept model inferred from the document title. The discourse about the time and duration of the spring occurrence of the moon's eclipse in the given year, 1659, establishes the discursive meaning of the quad. The document in question is a pamphlet authored by John Livie and published in 1659. It is entitled "The bloody almanack: or, Astrological predictions, and monethly observations, for the year, 1659. From the motions and configurations of the coelestial bodies, three great eclipses, one of the Sun, and two of the Moon: wherein is fore-told, the most eminent actions in Europe". Its EEBO-TCP identification number is A75144.

Further inspection of the full document title expands our apprehension of further discourses that are encompassed within the almanac. The title page lists the countries of Europe in 1659: "England, Scotland, France, Ireland, Denmark, Sweden, Poland, Flanders, Germany, And Hungary, Holland, etc." It elaborates

5. The EEBO metadata includes the subject designation: Almanacs, English; Astrology – Early works to 1800. It also includes the classification of these texts from the Short Title Catalogue (STC): "Astrology and cosmography Calendars, almanacs and prognostications".

too the particular prognostications regarding the wars and European battles anticipated, revolutions, and then, more prosaically, seasonal diseases, and laws and regulations that are likely to afflict the population. In a prefatory note to the reader, the author justifies the inclusion of 'bloody' in the titles of his annual almanacs, noting, "Because I writ it in Times, when the Heavens spoke little less to Mortals, but Blood, Trouble, and Commotions: I think the same consideration may excuse Me for the Title now". Indeed, a quick search yields the information that Livie's 1659 instantiation of the "Bloody Almanac" is among a highly eclectic collection of almanac pamphlets published between 1642 and 1666, characterised by titles that are highly political and religious sectarian in tenor.⁶

This additional encyclopaedic information further supports the reading of the quad's co-text. Zooming out from the immediate co-text reveals that the discourse within the almanac is not one-dimensional. Nested within the forecast of a spring moon eclipse is a sardonic comment on society. First Livie comments on the prevalence of diseases in the spring season owing to an excess of blood and choler, showing the contemporary common popular adherence to Galenic perspectives on humours and the human body. Then he offers a cryptic remark about the abuse of authority by "men pretending holiness at home" even as they are treated with courtesy abroad. This remark is completely opaque for the reader without access to the particular temporal, political and sectarian context in which this almanac would have been read.

The question is whether the deeper we dive into the local textual context, the more that narrowly specialist information is required to make sense of the co-text. The commentary accompanying the astrological prognostications is evident throughout Livie's almanac. The inspection of the zodiac each lunar month presages dire warnings of disease, tempests and destruction. In this particular case of the *day* quads, further exploration of the documents provides evidence that the astrological prognostications persist as the primary focus of the mid-seventeenth century almanacs, despite their liberal spicing up with acerbic comments on the morals and shortcomings of various parts of society. So regardless of the local

6. For example, "A bloody almanack foretelling many certaine predictions which shall come to passe this present yeare 1647. With a calculation concerning the time of the day of judgement, drawn out and published by that famous astrologer. The Lord Napier of Marcheston". (Anonymous. London. 1647. A1951.) "The bloody almanack: or An astrologically prediction of the most remarkable accidents which shall happen to the King, Parliament and city, as also in Scotland and Ireland, this yeer 1649". (Anonymous. London. 1649. A1953.) "The bloody almanack, or, Monethly observations and predictions, for the year of our Lord, 1654". (John Livie. London. 1654. A1921.) Alternate title: "Bloody almanack.; Monethly observations and predictions for the year of our Lord 1654".

textual context in the documents, the astrological concept model of the lunar calendar appears to hold for the *day* quads in Table 1.

3. Early hypothesis building and routes to relevant concept models

In this section, I analyse two quad sets to demonstrate the different routes by which we reach the most relevant candidates for concept models.⁷ In each case, we test whether and at what point early hypothesis building is rewarded. The additional information supplied by scrutinising the quads in contexts, including document metadata and the immediate co-texts in the documents themselves, enables us to query the integrity and stability of the concept models as well as the discursive roles they perform in the different documents. This step requires pragmatic analysis to generate the discursive concept.

Analysing the two sets indicates that there are shorter and longer routes of meaning assignment from quad data to concept models. The progress along these routes may be very rapid, as in the first case which consists of another set of highly frequently occurring quads around the node word *day*. However, the second case, quads around the node word *city*, indicates that mapping the quad-to-concept model might be an incremental process using encyclopaedic information and pragmatic inferencing (Carston, 2002; Blutner, 2004; Wilson and Sperber, 2004; Allan, 2007; Fitzmaurice, 2009). The cases demonstrate that concept model interpretation requires different amounts of effort to reach an acceptable level of specificity.

3.1 Day quads to Sabbath concept model

Table 2 consists of a set of twelve quads around the node word *day*. Their frequency of occurrence in EEBO-TCP ranges from almost 50,000 to just above 13,000. The most frequent one occurs in 1,266 documents; the least frequent in the table occurs in 470 documents.

The twelve quads in Table 2 share a common word set. *Sabbath* occurs in all but one quad and *lord* occurs in nine out of twelve. *Week* occurs three times, *name* (noun) twice, *rest* occurs twice as a noun and once as a verb, and *call* (verb) occurs twice. A reader whose heritage and experience are steeped in the western Judeo-Christian historical tradition will immediately discern a strong association between *day* and *Sabbath*. The prevalence of the trio *day-lord-sabbath*

7. Both sets of quads can be accessed from the lists of quads presented on the Linguistic DNA website; see: linguisticdna.org.

Table 2. Frequent quads around the node word *day*

Word	Word B	Word C	Word D	Freq.	MI score
day_n	keep_v	lord_n	sabbath_n	47,337	9.6327
day_n	holy_a	lord_n	sabbath_n	45,860	9.4704
day_n	call_vcc call v	lord_n	sabbath_n	45,401	9.8941
day_n	lord_n	sabbath_n	week_n	34,416	9.5460
day_n	lord_n	rest_n	sabbath_n	28,363	9.5659
day_n	lord_n	name_n	sabbath_n	21,840	9.8119
day_n	call_v	lord_n	week_n	21,829	9.5459
day_n	commandment_n	lord_n	sabbath_n	21,390	9.4448
day_n	call_v	sabbath_n	week_n	19,907	9.7126
day_n	lord_n	sabbath_n	sanctify_v	17,284	9.9033
day_n	rest_n	rest_v	sabbath_n	14,155	9.4943
day_n	call_v	name_n	sabbath_n	13,583	9.7500

contributes to this association. The discursive relationship inferred from the exceptionally strong association between the quad members amounts to the central precept captured in the fourth commandment: observing the Sabbath as a day for religious contemplation. The reader draws upon encyclopaedic knowledge and background, thus saturating the quads with familiar relevant contextual material to identify as the most relevant candidate the concept model of the Sabbath.

The mapping of meaning to this set of quads is a short and transparent process, triggered by the reader’s encyclopaedic knowledge and background. The robustness of this mapping can be tested by inspecting the metadata for the titles of the documents in which the quads occur, and then, if necessary, the occurrences themselves. The quads occur predominantly in catechisms, sermons and religious instruction tracts, and in political pamphlets regarding the observance of the Sabbath. We can consult examples in their original co-texts to examine the discursive role of the concept model. Examples (2), (3) and (4) illustrate the rhetorical functions performed by the concept model ‘Sabbath’. Example (2) is from a canonical document in which the fourth commandment is articulated and explicated through a question-and-answer format. The discursive relation between the quad members is the constitution of the commandment.

- (2) Quest. *The fourth Commandement.*
Answ. Exod. 20. v.9.10.11. Remember thou **keep'st** holy the **Sabbath Day**, sixe **dayes** shalt thou labour and doe all that thou hast to doe; but the seventh **day**

is the **Sabbath** of the **Lord** thy God: in it thou shalt doe no manner of worke, thou nor thy Son, nor thy Daughter, thy Man-servant, nor thy Maid-servant, thy Cattel, nor thy stranger that is within thy gates; for in sixe **dayes**, the **Lord** made Heaven and Earth, and all that in them is, and resteth the seventh **day**; therefore the **Lord** blessed the seventh **day**, and hallowed it.

(A60312. *A new catichisme for ever: or, The mite of a minister Cast into a catechisme, for the preparation of his people, before they receive the supper of the Lord.* By Joshua Siston, minister at Grauby, in the vale of Bever. Siston, Joshua, b. 1603 or 4. 1650 (1650). Wing S3909A; ESTC R221499.)

Example (3) is also from a work of religious instruction. However, this text reviews the underpinning biblical genealogy of the contemporary mid-seventeenth century Anglican tradition of observing the Sabbath (Sunday rather than Saturday). The author, Isaac Penington, cites the precedence for keeping the true Sabbath holy, distinguishing between the “day of redemption” or the resurrection of Christ which Christians commemorate by observing Sundays as the holy day, and Moses’ designated day of rest (the Jewish Sabbath). The density with which the quad members cluster and the occurrence of other salient lemmas in this set of quads here, including *name*, *keep* and *holy*, support the weight of the Sabbath concept model in this explicative context.

- (3) For he that hath not received the Law of the Spirit of life in Christ Jesus, knoweth not the Lord of life to be the onely true God, but maketh Images in his mind, and taketh his Name in vain, not feeling the living power thereof, nor can keep his **Sabbath** (ceasing from sin, forbearing his own works, his own willing and running, and entring into the rest of the Gospell) for there is but one day of rest **holy** to the **Lord**; all the **Sabbaths** of the Law were but signes of it, having but a significative or representative **holynesse**; but the **day** of redemption which the **Lord** hath made, *Psal.* 118. 24. (in which his redeemed rejoyce, and **rest** to him) that hath the true **holynesse**. This was it which came by Christ, the other came by *Moses, Joh. 1. 17.* *Moses* his family or children were to keep that **day** (that was the **day** for the servants, who were to be exercised under the shadows) but the believers are to keep this **day** in the Spirit, to enter into this **rest** by the faith.

(A90390. *An epistle to all such as observe the seventh-day of the week for a sabbath to the Lord.* Penington, Isaac, 1616–1679. 1660 (1660). Wing P1164; ESTC R229116)

The quad *day-call-lord-Sabbath*, one of the most frequent in Table 2, includes the prevalent trio *day-lord-Sabbath*. In Example (4), the quad co-occurs with other quad set lemmas; for example, *rest*, *keep* and *holy*. This document is a political tract arguing for a particular spiritual interpretation of the fourth commandment. The author cites sources of Christian authority, like St Jerome, St Augustine

and Tertullian, to support the argument that if the faithful observe the Sabbath, the holy day, through spiritual rather than physical rest, then every day should be as holy as the designated Sabbath: Sunday. He opposes the obedient and restrained Christian who works religiously and conscientiously on a Sunday to the evil "Preachers of Antichrist" who forbid men to work on the Sabbath.

- (4) That the true Observation of the Sabbath consisteth not only in abstaining from bodily labors; And that to a Christian man every day is the Sabbath, and not only the Seventh day. *Hierome* saith, Therefore be certain **dayes** assigned, that we should come together; not that that **day** in the which we come together is **holier** then another, but all **dayes** be alike and equall. And Christ is not alonely crucified in Parasceden, and risen only on the Sunday, but the **day** of Resurrection is alwayes, and alwayes may we eat of the **Lords** Flesh, &c. *Augustine* saith, We must observe the Sabbath **day**, not that we should reckon our selves not to labor, but that all things that we do work well, must have an intention to the everlasting **rest**. Wherefore we must observe the **holy day**, not by corporall idleness, and unto the Letter, but spiritually must we **rest** from vices and concupiscences. Wherefore among all the ten Commandements, that of the Sabbath **day** is alonely commanded to be figuratively observed, &c. *Tertullian* saith, The carnall Circumcision is put away, and extincted at his time. So likewise the observation of the Sabbath **day** is declared to be for a time, for we must **keep** the Sabbath **day**, not alonely the Seventh **day**, but at all times, as *Isai.* saith, &c. *Augustine* saith, It is come unto me, that certain men, which be of an evill mind, have sowed certain evill things among you, and contrary to the **holy** Faith, so that they do forbid that men should work on the Sabbath **day**.

(A82315. *The doctrine of the Sabbath, as it hath been believed and taught, by ancient and eminent Christians, collected word for word out of their own writings, and now tendred to the consideration of all the godly, especially to direct them to the Parliament, to direct them in their intended Act, for the due and strict observation of the Lords Day.* / By a friend to truth, and to the present powers of this Common-wealth, in the way of truth. Dell, William, d. 1664. 1650 (1650) Wing D922; Thomason E597 14; ESTC R206297)

The concept model in Example (4) is a vehicle for promoting a particular version of organised religion, which permits men to conduct necessary business on the Sabbath. In this version, the faithful are urged to carry out this virtual and virtuous practice every day. In contrast, if the obligation to observe the Sabbath is construed literally as withdrawing labour on one day, the Sabbath, that practice is characterised as anti-Christian, and thus not to be tolerated. The hermeneutic work required of the reader to understand the situation of the concept model within the discourse involves careful close reading, and constant assessment and

reassessment of the discursive meaning by saturating the linguistic form with an expanding context. This example represents the partisan sectarian attempt to advance a particular interpretation of the fourth commandment to cast aspersions on established interpretations. Here, the concept model is recruited for particular rhetorical purposes – namely, to shame readers who do not share the views of the author.

Examples (2) to (4) indicate that the concept model's integrity survives the different discursive settings for which it is recruited. The early hypothesis building of a concept model around the fourth commandment is confirmed incrementally through the steps of consulting the metadata of the documents in which the quads occur, and then the original co-texts of the quads.

3.2 City quads to taxation concept model

The second case (mapping meaning onto the quads around the node word *city*) requires an iterative process of hypothesis building that draws upon more, and more detailed encyclopaedic information to reach the most relevant hypothetical concept model. It is not possible to produce a feasible concept model without highly specific historical, generic and temporal information that saturates the context in which it can be interpreted. The resulting concept model is highly stable owing to its technical affinity with the dominant genre of the ordinance. Table 3 contains a set of quads around *city* that share an overwhelmingly common word stock.

Table 3. Frequent quads around the node word *city*

Word A	Word B	Word C	Word D	Freq.	MI score
city_n	county_n	pound_n	sum_n	53,256	13.2058
city_n	county_n	pound_n	shilling_n	34,921	14.4916
city_n	county_n	shilling_n	sum_n	31,425	12.7465
city_n	pound_n	shilling_n	sum_n	29,780	12.5631
city_n	county_n	penny_n	pound_n	29,388	12.6330
city_n	county_n	penny_n	sum_n	28,155	12.7686
city_n	penny_n	pound_n	sum_n	26,410	12.5513
city_n	county_n	penny_n	shilling_n	23,253	14.0219
city_n	penny_n	pound_n	shilling_n	22,719	13.6986
city_n	penny_n	shilling_n	sum_n	20,709	12.1572

The most common words around the node word *city* are *county*, *penny*, *shilling* and *sum* (six quads), and *pound* (five quads). If a user encountered these quads in isolation, without any knowledge of the corpus from which they come, initial hypothesis building might be impossible owing to the difficulty of calculating the connections between their members. In isolation, each item may be vague and underspecified; *city*, for example, carries some ambiguity as a physical geographical urban entity or as an organisation. Similarly, *pound* is a unit of measurement; but in isolation, it is hard to determine what it measures (weight, currency) or whether it is literal or figurative (pound of cheese, “pound of flesh”). Other terms might be hard to construe; one is *county*: an administrative unit or division peculiar to the British Isles and its colonies – an Anglo-Norman term introduced in the fourteenth century. This knowledge is either integral to a reader’s cultural background or acquired by looking it up.

How can we build a hypothesis at this point about the association between *city* and the other members of the quad? Closer inspection of the quads yields two lexical fields; *city* and *county*; and *penny*, *shilling* and *pound*. We might identify the connection between *city* and *county* as administrative because cities are located in counties and a city like York might serve, historically, as an administrative centre for a county like Yorkshire. The second set consists of expressions that refer to units of a particular currency. The term *sum* refers to a quantity of money and, thus, easily connects with the larger word set. Both word sets are highly localised and specific to British history, culture and experience, as they apply to organising the administration of a country and managing its currency.

Hypothesis building points to linking money with counties and cities; we might assume simply that cities need and make money. To resolve the matter, we draw upon our encyclopaedic knowledge and experience. (For example, the City of London is nicknamed “Londongrad” as the global capital city for money laundering, particularly by Russian oligarchs.) However, because the quads in Table 3 occur in the EEBO-TCP universe of discourse, they are highly temporally and situationally specific to the England and the British Isles of the Early Modern period. This specific context prompts the hypothesis that what links the city and money is raising or paying taxes or seeking to levy funds from the citizens of a city or county. Accordingly, the concept model that best maps onto the *city-county-pound-sum* quad, and the other quads in Table 3, is the notion of raising money from citizens.

To test the robustness of the concept model, we inspect the metadata of the documents in which the quads occur. The top quad occurs in sixty-seven documents whose titles indicate that they are ordinances, orders and declarations, acts of Parliament, or proclamations that demand or authorise the raising of funds for armies, ships or wars, or for the King or for the Protector (depending on the date).

The elaborate descriptive details in the document titles in Examples (5a) to (5d) are clear clues to the documents' contents.

The document title in Example (5a) directs "the knights, citizens and burgesses of the severall counties, cities, and boroughs of this kingdom" to co-operate with the commissioners appointed to assess and levy the funds to be raised through an Act of Parliament. The information about raising funds across England and Wales to pay for the peace process and disarming and disbanding of armies in the England–Scotland conflict confirms the notion of raising money from citizens as the concept model.

- (5a) It is this day ordered by the House of Commons, that the knights, citizens, and burgesses of the severall counties, cities, and boroughs of this kingdom, and the dominion of Wales ... shall forthwith signifie unto the commissioners appointed for the assessing and leavying of all summes of money to be raised by an act of Parliament intituled, An act for the speedie provision of money for disbanding the armies and settling the peace between the two kingdoms of England and Scotland. 1641. (A38239)
- (5b) An ordinance of the Lords and Commons assembled in Parliament, for the maintaining of the forces of the seven associated counties, under the command of Edward Earl of Manchester. By a weekly payment upon the said associated counties; to begin the first day of September, and to continue for foure moneths next ensuing. : Septemb. 26. 1644. / Ordered by the Commons in Parliament, that this ordinance be forthwith printed & published. ; H. Elsynge, cler. Parl. D. Com. (A83209)
- (5c) An ordinance of the Lords & Commons assembled in Parliament, for the raising of moneys to be employed towards the maintenance of forces within this kingdom, under the command of Sir Thomas Fairfax knight. And for the speedy transporting of and paying the forces for carrying on the War of Ireland. / 7 Julii, 1647. Ordered by the Commons assembled in Parliament, that Mr. Husband do forthwith print six thousand of the ordinance for raising moneys for maintenance of the army under the command of Sir Thomas Fairfax knight. ; H: Elsynge, cler. Parl. D. Com.; Laws, etc. (A83241)
- (5d) By the King and Queen, a proclamation for prolonging and appointing the time for the first general meeting of their majesties commissioners for executing the Act of Parliament lately made for granting to their Majesties an aid of twelve pence in the pound for one year, and for authorising and empowering the respective commissioners to proceed and act accordingly. William R.; Proclamations. 1689-09-02. (A66307)

The document titles in Examples (5b) and (5c) are for ordinances dating from 1644 and 1647, respectively, announcing the raising of funds to support forces

under the command of the Earl of Manchester involved in the English civil wars. The 1644 document specifies that the funds will be collected by a weekly payment to last for four months. The temporally specific, situated content of this title underlines the concept model of raising money from citizens as tax. Similarly, the title in Example (5c) announces the raising of money to support forces under Sir Thomas Fairfax's command to prosecute a war with Ireland. These examples illustrate that the document titles alone carry sufficient information to confirm the concept model. It is possible that it is the particular genre of the ordinance or official order (which is quite brief and carries highly time- and context-sensitive information) that enables the easy confirmation of the concept model.

Of course, digging further into these documents and mapping them to a specific timeline yields more detailed information about the period and the political conditions in which they were issued than their individual titles can. The titles provide a clue about how the Parliamentary commanders managed to develop a suite of instruments for raising taxes and financing battles to wage war against Charles I's Royalists in the English civil war, between 1642 and 1648. Example (5d) dates from 1689, the year in which William of Orange and his wife, Mary, were crowned as joint sovereigns. This ordinance was issued in September 1689 to implement a Parliamentary act to secure financial support for the King and Queen for one year. The reason for selecting this ordinance in addition to those in Examples (5a) to (5c) is to underline how far government finance, whether monarchy or commonwealth, depended deeply upon the ability of the English Parliament to raise funds through taxes and levies upon its citizens, in cities and counties across the realm.

Clearly, the detailed background of the ordinances of the period yields material that validates the hypothesis that the concept model is about a kind of taxation. In each example, the concept model is embedded in the very specific, narrow details of place, time, events and actors spelt out in each ordinance. Nevertheless, the taxation concept model easily survives its highly specific textual embedding. One factor that ensures the concept model's stability is the stable function of the ordinance genre or text-type as the primary vehicle for communicating the legal framework and method for implementing the taxation over time.

The question is whether analysis of the metadata of the document titles and the detail of the documents themselves is necessary to map meaning onto the quad data to generate the concept model. Two factors are key to the interpretation of the quads around *city* in Table 3. The first is access to a set of quads that share a common word-stock. However, the properties of under-specification and vagueness that mark individual lexical items in isolation present obstacles to early hypothesis building. In addition, some lexical items in a quad may be unconstructable without special knowledge of their potential relevance to a particular domain

(for example, English currency or British administrative divisions). The domain specificity and thus interpretative obscurity of individual items militates against initial hypothesis building without easily accessible encyclopaedic information to make sense of the items, both in isolation and in association. The second factor, therefore, is necessary access to the relevant encyclopaedic knowledge or information to build the initial hypothesis. The next step – inspection of the metadata of the document titles in which the putative concept model occurs – permits validation of the hypothesis where access to encyclopaedic information is present. Where early access to encyclopaedic information is impossible, the inspection of the metadata is necessary for the concept model building. In this context, further examination of the quad in its original co-text in the documents provides a means of confirming the hypothetical concept model.

4. Discussion and observations

The meaning mapping procedures rehearsed to reach the “Sabbath” and “taxation” concept models illustrate two ways of ascertaining the value of early hypothesis building in the process of assigning meaning to quads as concept models. They also illustrate how the process takes different routes, some short and simple, and some more complex and incremental.

The “Sabbath” concept model resulted from mapping meaning to a set of quads around the node word *day*. Mapping meaning in this case is a short and transparent process, triggered by the reader’s encyclopaedic knowledge and background. The robustness of the mapping process was tested by inspecting the metadata for the titles of the documents in which the quads occur, and that test confirmed the analytical value of early hypothesis building. The next step was to scrutinise the occurrences of the quads in the documents themselves to see whether the particular discourse alters the concept model to produce a different, perhaps particularly context-bound discursive meaning. The examination of the quads in their original co-texts indicates that whether the author is using the “Sabbath” concept model to explain a point of catechism or to argue against a particular religious observation practice, it survives the rhetorical intentions for which it is deployed. This means that, regardless of the genre or register to which the document belongs, the concept model remains intact.

The hypothetical meaning assignment of the “taxation” concept model to the quad set around *city* requires more work to ascertain the referential specificity of the quad members and the nature of the strong association between them. An iterative process of hypothesis building was required at the level of the individual lexical items as well as at the pair, trio and quad levels. Encyclopaedic information

played a crucial role in providing sufficient detail to hypothesise the most relevant candidate meanings for the quads and to test their robustness.

What is striking about both of the cases is that the concept models are demonstrably highly stable in the sense that they remain intact regardless of the discursive contexts in which they occur in the EEBO-TCP documents. What is equally striking is that the derivation of the concept models depends upon retrieving encyclopaedic knowledge. However, while meaning assignment to the *day* quads to reach the “Sabbath” concept model involved a transparent and simple route, reaching the “taxation” concept model required a set of procedures. In particular, the interpretation of the “taxation” concept model required repeated saturation of the quads around *city*, with detailed and specific information as context for hypothesis building. In contrast, the “Sabbath” concept model required less, and less iterative pragmatic work to reach the most relevant meaning for the quad set.

5. Conclusion

The stable concept models examined in this study contrast with the volatile quads explored in Fitzmaurice and Mehl (2022) which require iterative semantic–pragmatic work drawing upon encyclopaedic knowledge and the repeated application of pragmatic routines, including the resolution of vagueness, specification, saturation and inferencing. The process of early hypothetical meaning assignment explored in this study suggests that highly stable concept models can be mapped onto some quad sets through a short route requiring minimal pragmatic work, in addition to the reliance on encyclopaedic knowledge. The implications of this discovery for users of the Linguistic DNA Concept Modeller are manifold. If the step from quad outputs to stable concept model consists of mapping with minimal effort, then users will be able to easily access and interpret the outputs of the concept modeller. They will also be aided in understanding the work being carried out by those concept models in early modern discourses and the stability of the concept models.

The Linguistic DNA project work has so far demonstrated that the outputs of the Concept Modeller require different approaches to the assignment of meaning to them as concept models and to the hermeneutic work of construing their meanings in discourse. The work reported here complements the finding (Fitzmaurice, forthcoming) that regardless of the simplicity or complexity of assigning meaning to quad data, a concept model can retain its conceptual integrity across the different discourses in which it occurs. The analysis also provides strong evidence that the stability of the concept models examined can be interpreted as stability across time as well as across discourses even though

they are recruited for very particular rhetorical purposes in very particular early modern discourses.

Funding

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

References

- Allan, Keith. 2007. "The Pragmatics of Connotation". *Journal of Pragmatics* 39 (2007): 1047–1057. <https://doi.org/10.1016/j.pragma.2006.08.004>
- Blutner, Reinhard. 2004. "Pragmatics and the Lexicon". In Laurence R. Horn and Gregory Ward (eds), *Handbook of Pragmatics*, 488–514. Blackwell. Oxford.
- Burns, Philip R. 2013. "MorphAdorner v1: A Java Library for the Morphological Adornment of English Language Texts". Evanston, Illinois: Northwestern University. Accessed 29 November 2016 at: <https://morphadorner.northwestern.edu/morphadorner/download/morphadorner.pdf>
- Carston, Robyn. 2002. *Thoughts and Utterances: The Pragmatics of Explicit Communication*. Oxford: Blackwell. <https://doi.org/10.1002/9780470754603>
- Fitzmaurice, Susan. 2009. "The Sociopragmatics of a Lovers' Spat: The Case of The Eighteenth-century Courtship Letters of Mary Pierrepont and Edward Wortley". *Journal of Historical Pragmatics* 10 (2): 215–237. Reprinted in Jonathan Culpeper (ed.), *Historical Sociopragmatics*, 37–61. Amsterdam: John Benjamins.
- Fitzmaurice, Susan. Forthcoming. "From Constellations to Discursive Concepts; or: The Historical Pragmatic Construction of Meaning in Early Modern English". *Transactions of the Philological Society*.
- Fitzmaurice, Susan and Seth Mehl. 2022. "Volatile Concepts: Analysing Discursive Change through Underspecification in Co-occurrence Quads". *IJCL-22005 (International Journal of Corpus Linguistics)*.
- Fitzmaurice, Susan, Justyna A. Robinson, Marc Alexander, Iona C. Hine, Seth Mehl and Fraser Dallachy. 2017. "Linguistic DNA: Investigating Conceptual Change in Early Modern English Discourse". *Studia Neophilologica* 89 (Supp. 1): 21–38. <https://doi.org/10.1080/00393274.2017.1333891>
- Mehl, Seth. 2019. "Measuring Lexical Co-occurrence Statistics Against a Part-of-Speech Baseline". In Hanna Parviainen, Mark Kaunisto and Päivi Pahta (eds), *Corpus Approaches into World Englishes and Language Contrasts*, VARIENG e-series vol. 20. Helsinki: VARIENG. Accessed 1 September 2020 at: <http://www.helsinki.fi/varieng/series/volumes/20/mehl/>
- Mehl, Seth. Forthcoming. "Discursive Quads: New Kinds of Lexical Co-occurrence Data with Linguistic Concept Modelling". *Transactions of the Philological Society*.
- Wilson, Dan and Deidre Sperber. 2004. "Relevance Theory". In Laurence R. Horn and Gregory Ward (eds), *The Handbook of Pragmatics*, 607–632. London: Blackwell Publishing.

Address for correspondence

Susan Fitzmaurice
Humanities Research Institute
University of Sheffield
Gell Street
Sheffield, S3 7QY
United Kingdom
s.fitzmaurice@sheffield.ac.uk

Publication history

Date received: 1 March 2022
Date accepted: 5 May 2022