

Aatu Liimatta
aatu.liimatta@helsinki.fi

DOI: 10.4312/linguistica.63.1-2.353-374



Jani Marjanen
jani.marjanen@helsinki.fi

Tuuli Tahko
tuuli.tahko@alumni.helsinki.fi

Mikko Tolonen
mikko.tolonen@helsinki.fi

Tanja Säily
tanja.saily@helsinki.fi

University of Helsinki

DIMENSIONS OF INCOMING ECONOMIC VOCABULARY IN EIGHTEENTH-CENTURY BRITAIN

1 INTRODUCTION

The intellectual history of the eighteenth century, characterized by the emergence of the Enlightenment and its platform of modernity, has been increasingly linked to economic improvement in contemporary discourse (see especially Robertson 2005). Some scholars also use descriptions such as “economic enlightenment” (Popplow 2010) or “agricultural enlightenment” (Jones 2016) to foreground the whole period as particularly economic or agricultural. The crux lies not in a conventional definition of Enlightenment’s nature, but rather in comprehending the practical evolution of eighteenth-century thought. This interest in the economy is reflected in changes in eighteenth-century vocabulary, which is often pointed out in scholarship (Shovlin 2006: 2–5, McIntosh 2020: 163–165), but not studied rigorously in terms of language use, and only mentioned briefly in e.g. overviews of English historical lexis (Nevalainen 1999). Biber/Finegan (1997) have shown that from the eighteenth century onwards, English registers diverged in their language use: speech-based and popular written registers like fiction became more colloquial, whereas expository ‘specialized’ registers like academic texts became more literate in style.¹ As the English language underwent significant functional expansion stimulated by various socio-cultural changes during the eighteenth

1 “Literate” in this context is to be understood as opposed to “oral” or “conversational”; according to Biber & Finegan (1997: 260), literate registers are “characterized by careful, sustained production circumstances; informational communicative purposes; and minimal interactivity”.

century (Culpeper/Nevala 2012), and considering the contemporary interest in the economy, we hypothesize that this expansion drove the spread of new economic vocabulary both in terms of its frequency of use but also in its application in both concrete and increasingly abstract contexts. The main emphasis of our paper is to utilise a data-driven approach to draw conclusions about economic vocabulary. While we have identified various potential factors that may have influenced its expansion, our goal in this study is to primarily observe and analyze the data, rather than getting involved in a discussion about the underlying reasons. This is particularly the case with regard to potential underlying economic developments that may or may not have influenced the linguistic expressions. While they are interesting, the correlations are difficult to chart due to the difficulty in comparing linguistic and economic data.

In combining linguistic and historical expertise, we make use of novel computational methods and draw on the massive database of *Eighteenth Century Collections Online* (ECCO) to trace incoming economic vocabulary in British publications dealing with economic matters in the widest sense (about ECCO, see Tolonen et al. 2022). As the starting point of our analysis of economic vocabulary, we focus on the words classified by the *Historical Thesaurus of the Oxford English Dictionary* under the category ‘trade and finance’. In order to identify incoming economic vocabulary in particular, we apply a novel method, the Proportional Change Rate (PCR; Nevalainen et al. 2020), a measure developed for identifying periods of rapid and slower linguistic change as well as changing linguistic items. Furthermore, in order to gain more insight into the lexical functions and underlying discourses of the incoming vocabulary, we make use of an approach inspired by the multi-dimensional method of register analysis (MDA; Biber 1988) in which we apply the multi-dimensional method to the economic content words instead of the functional features of typical MDA register analyses; this is similar to other recent developments in applying a multi-dimensional approach on content words (cf. e.g. Zuppari/Berber Sardinha 2020). We also combine our textual data with metadata from the *English Short-Title Catalogue* (ESTC), which has been harmonized and augmented by the Helsinki Computational History Group (Lahti et al. 2019).

This novel approach does not come without problems, however. As an example, the digitized texts in ECCO have been automatically converted from scanned images into plain text, and this optical character recognition is riddled with errors impacting both precision and recall (Hill/Hengchen 2019). This pilot study aims to be a step towards a systematic quantitative assessment of economic language use based on these large-scale collections, taking a critical view on the prevailing challenges in such analysis. With the combined expertise of the linguists and historians in our group, we will shed new light on the development of economic discourse by producing a description that is better couched in historical linguistics than earlier claims of an “economic enlightenment”, and pave the way for further research utilizing historical language data in the digital humanities.

2 BACKGROUND

The increased written use of vernacular languages (as opposed to the cosmopolitan Latin) was an uneven process, in the sense that some local languages became print languages earlier than others and some registers or genres were more prone to vernacularization. In the medieval period, devotional literature comprised the majority of vernacular manuscripts in Europe, and the vernacular sermon, for instance, required the learned to communicate with lay people (Crossgrove 2000, Muessig 2010: 267). The fields of science, technology, and medicine also had a significant vernacular readership already in the Middle Ages (Crossgrove 2000). Religion and medicine, in particular, were areas of life where the educated mixed with the uneducated, which may have in part prompted vernacularization in these domains. Similarly in the eighteenth century, the insight of needing to reach a larger proportion of the population was particularly present in discussions about economic improvement. As states were increasingly seen as being in economic competition with one another (Hont 2005), the application of economic reform demanded bridging academic knowledge with the everyday practices of civic organizations and concrete economic policy. Economic matters became at the same time more specialized as they became the object of increased academic inquiry, and accessible to a broader readership as key actors increasingly saw a need to not only produce new knowledge, but also disseminate it to practitioners outside the sphere of academia.

The specialization of this knowledge is best visible in several chairs containing the word *economy* in their titles being established at European universities during the course of the eighteenth century (Magnusson 1992: 249–257) and an accentuated theoretical debate with treatises reasons and nature of prosperity in different European countries. Some volumes, like Plumard de Dangeul's *Remarques sur les avantages et desavantages de la France et de la Grand-Bretagne* (1754), very concretely highlighted the economic rivalry between Great Britain and France, making them of great interest to the reading public, but most probably served to disseminate economic vocabulary. As debates about the economy crossed borders, so did the terminology, which is perhaps reflected in the large portion of French borrowings in the new eighteenth-century economic vocabulary in English.

In more practical terms, so-called economic or improvement societies that proliferated all over Europe in the eighteenth century (Stapelbroek/Marjanen 2012) ventured into new avenues to put new knowledge, instruments and models into practice. The Honourable Society of Improvers founded in Scotland in 1723, which is generally regarded as the first of such societies, not only gathered information for discussion among its members, but actively published accounts of how to improve crops and manage farms in the newspapers (Bonnyman 2012: 37–38). Often written communication was not enough. In disseminating knowledge of new technology, The Dublin Society for the Improvement of Husbandry, Agriculture and other Useful Arts (founded in 1731), not only distributed models and drawings of ploughs to be used with different crops and types of soil, but also started promoting ploughing matches to promote efficient new agricultural methods (Livesey 2012: 69).

All this required both new vocabulary and new publication channels through which this new terminology could reach newly relevant strata in society. Overall, the role of the dictionary as a platform for lexical innovation becomes evident from Chambers onwards.. For instance, encyclopaedia entries on economic matters multiplied from Diderot and d'Alembert's *Encyclopédie* from the mid-century to Panckoucke's *Encyclopédie Méthodique* published toward the end of the century (Shovlin 2006: 2–5). Many seventeenth century or early eighteenth century neologisms, such as *political economy*, *commercial*, *consumption*, and *industry*, spread in their use during the course of the eighteenth century (McIntosh 2020: 163–165). Many of these terms expanded the horizon of economic thinking from the practical to the more theoretical and further from local circumstances to a national (or even international) perspective. The remarkable proliferation of eighteenth-century dictionaries listed in the English Short Title Catalogue is striking.

Economic vocabulary also penetrated more popular styles and registers. This can perhaps be seen in the orthography of some words, *economy* being the most obvious case. The form 'economy' overtook the forms 'oeconomy' and 'æconomy' by the end of the eighteenth century, marking a shift from spelling the word in the "classical" to a modern "English" way. Such orthographic changes stemmed from the gradual creation of local standards, printing technology and beliefs of what would serve the public best. In the 1759 book *The abecedarian* John Yeomans argued for dropping the ligature spelling for the word 'economy' and other foreign words as "we refuse to gesticulate the modes of spelling by any nations upon earth" (Yeomans 1759: 48). While some other suggestions, such as writing 'se' instead of 'sea', 'te' instead of 'tea' or 'æful' instead of 'awful', did not catch on, the overall aim of remedying "the present condition of our vulgar tongue" and making "learning easy" (Yeomans 1759: 15) was a driving force in reforming spelling.

The vernacularization of economic language can be seen as taking place in the intersection of new fora, new ideas about how to communicate economic knowledge, and the spread of both new and established economic vocabulary. Our focus lies on the last one of these, but our analysis is grounded in the idea that linguistic innovation was couched in the institutions, publication practices and the interplay between different registers that helped produce new lexical items as well as make them part of everyday language.

3 MATERIAL AND METHODS

3.1 Material

The dataset we employ is based on texts from the *Eighteenth Century Collections Online* (ECCO), which contains full texts of some 200,000 documents as computer-readable texts produced by optical character recognition (OCR). These documents have been linked to corresponding metadata from the *English Short Title Catalogue* (ESTC) by the Helsinki Computational History Group (COMHIS); this metadata has then been further refined and enriched (Tolonen *et al.* 2022). As our focus is on the spread and change of economic language, we have chosen to further focus on a specific subset of documents: records labelled in the ESTC as belonging to the Goldsmiths'-Kress

Library of Economic Literature (GKL), a microfilm collection of economic literature in a broad understanding of “economic” (Whitten 1978).

When building our dataset based on this collection, we only included the earliest edition of any given work, since any reprints of older works will likely include linguistic features characteristic of their original period of publication, and thus may confuse diachronic analyses. Taking also into account the fact that not every ESTC record is represented in ECCO, we ended up with a dataset of roughly 5,000 ESTC records linked to ECCO documents, down from the full 11,000 ESTC record GKL set. Because an ESTC record can map onto multiple documents in ECCO (e.g. in the case of multi-volume works), this set of roughly 5,000 ESTC records corresponds to a set of about 5,200 ECCO documents. Figure 1 shows the number of documents over time divided by type. We can see that the number of pamphlets included in our data is particularly high in the early decades of the century. This is related to the 1707 Union of England and Scotland, which also led to an increase in pamphleteering on economic topics. Although the document count of pamphlets in GKL in the 1710s is high, this did not have an impact on our analysis due to both the low word-count of the pamphlets and, more importantly, the fact that the Union debate did not include the incoming economic vocabulary that turned out to be central for this study.

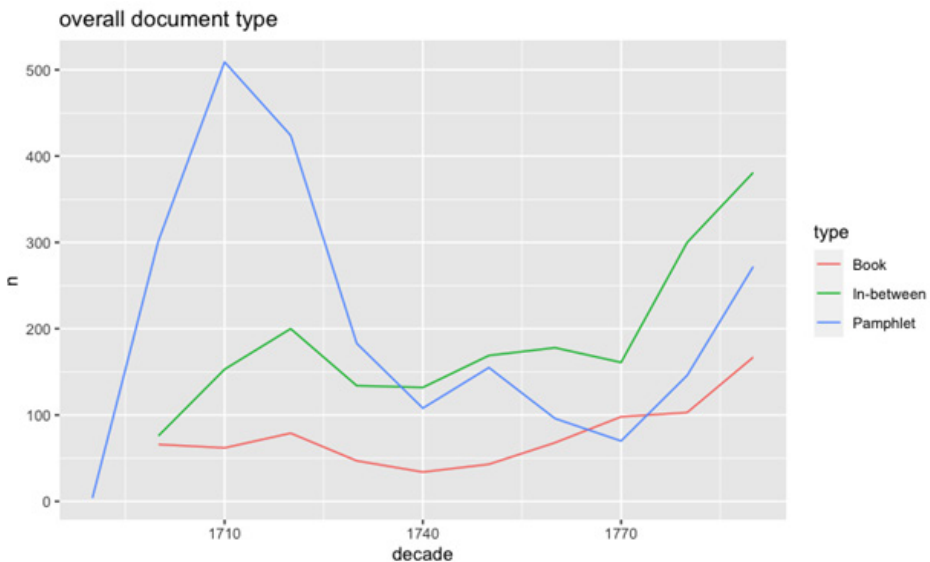


Figure 1. Number of ECCO documents in the GKL over time, divided by document type (pamphlet = <32 pages, in-between = 32–128 pages, book = >128 pages)²

2 About these different document types, see Mikko Tolonen, Eetu Mäkelä, and Leo Lahti, “Supplementary Information to The Anatomy of Eighteenth Century Collections Online (ECCO) article,” *Supplement to ‘Anatomy of ECCO,’* June 2022. <https://doi.org/10.5281/zenodo.6683914>

As the starting point of our analysis of economic lexis, we take the ‘trade and finance’ section of the *Historical Thesaurus of the Oxford English Dictionary* (HT). Edited by a team led by Professor Christian Kay at the University of Glasgow, the HT describes the semantic development of English based on the second edition of the *Oxford English Dictionary* (OED) and *A Thesaurus of Old English*. It rearranges the OED by meaning into hierarchical semantic classes under the top-level categories of ‘the mind’, ‘society’ and ‘the external world’. Out of the 800,000 words and meanings included in the thesaurus, 14,989 distinct lexical items can be found under the category of ‘society > trade and finance’.

Thanks to an agreement with Oxford University Press, we have access to local XML versions of the OED and HT. This enables us to easily retrieve not only a list of all words belonging to a HT category but also the lexicographical metadata associated with each word, such as etymology type (e.g. borrowing or derivative), etymon language (e.g. Latin or English), part of speech, year of first attestation, and the specific HT class of the word (e.g. ‘society > trade and finance > illegal or immoral trading > trade in (goods) illegally or immorally > smuggle’).

Using the OED as a source for first attestation dates in particular is not unproblematic (*cf.* Durkin 2002). Work on the third edition is ongoing, and there are still entries in the dictionary that remain essentially unchanged since the first edition, prepared in the late nineteenth and early twentieth centuries. It is very probable that some of the entries have been antedated since March 2019, when we acquired the data, or will be antedated in the near future. Nevertheless, our usage of the OED is not dependent on the dating being absolutely correct. We are interested in economic lexis, whether old or new, that was extant in the eighteenth century and that increased in frequency during that century in ECCO. To make sure that we are not missing words with economic senses that were extant in the eighteenth century but that the OED dates somewhat later, we take all words in the HT ‘trade and finance’ category with an OED first attestation date starting from Old English and ending in 1850. The cutoff of 1850 rather than 1800 accounts for possible antedatings to our OED data. In our investigation of specific lexical items associated with the dimensions of incoming economic vocabulary identified in our analysis (4.1 below), we retrieved the first attestation dates from OED Online in spring 2022. While the same caveats apply, the datings are used as indicative of the relative age of the words, particularly across the different dimensions, and our main findings regarding the dimensions do not rely on exact dates.³

The OCR of the ECCO data used is not perfect and causes some self-evident problems in matching strings of characters. In general, the distribution of errors is relatively even so that the results of most of our quantitative analyses are not affected by poor OCR. There are, however, some exceptions to this: the use of the long ‘s’ in eighteenth-century documents, as well as the use of ligatures, both of which often include errors in the machine-readable version (Hill/Hengchen 2019). We manually inspected the OCR

3 In analyses which focus more heavily on the dates of introduction of new lexical items, the OED data could be supplemented by e.g. lexical lookups of the full ECCO dataset and other datasets such as the British Newspaper Archive.

quality of the spelling variants ‘oeconomy’, ‘æconomy’ and ‘economy’ in a selection of texts, and could confirm that about a third of the machine-readable strings ‘economy’ were, in fact, false renderings of the form ‘æconomy’. Regardless, our manual inspection shows that the spelling variant ‘economy’ did become more common in the last decades of the eighteenth century, but any detailed quantitative assessment of spelling variation would not be reliable without manual inspection due to ligatures being exceptionally prone to errors. Based on this inspection, we also note that results with regard to words with the letter ‘s’ potentially include a higher rate of errors.

3.2 Methods

In order to study the changes in economic vocabulary, we make use of a two-phase methodology. In the first phase, we identify incoming economic vocabulary; in the second phase, we identify underlying discourses driving the rise of some of these words.

3.2.1 Incoming economic vocabulary

We start our analysis of economic vocabulary with the set of lexical items classified under ‘trade and finance’ in the HT. As our focus is on the eighteenth century, we further limit this set to those items which are first attested in the ‘trade and finance’ category before the year 1850. We set this limit to avoid including items which were used with other meanings in the eighteenth century but which only later developed an economic meaning, but offset the limit by 50 years because it is possible there may be antedatings in our dataset of the first attestations recorded in the OED, as discussed in 3.1 above. We also exclude items whose entries contain parenthetical remarks or alternative constructions, which would not appear in that form in running text. After the set of items was determined, the frequency of each of the 7,079 items remaining in this set was calculated by decade by dividing the total number of occurrences of the item in each decade of our data by the total number of characters⁴ in the OCR versions of the documents.

To identify items which rise in frequency relatively consistently throughout the century, we calculated the *proportional change rate* (PCR; Nevalainen *et al.* 2020) of each of the items across the decades. Originally devised to help identify time periods with the highest rate of change, the PCR describes how much of the total change in the frequency of a lexical item happens between each pair of successive time periods. In order to focus on incoming items with relatively constant growth, we excluded all items which had a PCR higher than 10% during any period of negative growth, that is to say, items which had at least one period of decreasing frequency which constituted over 10% of their total frequency changes. To avoid spurious results and statistical noise, we further excluded all items which had fewer than 100 total occurrences throughout our dataset, as well as items which did not appear in at least two decades. In the end, the set of lexical items of potential interest included 73 items. Figure 2 shows that the

4 Even though using the word count of the documents as the normalization base would be preferable, we use the character count instead to mitigate the effects of the questionable OCR quality, which makes it difficult to determine the word count of a document.

increase in frequency happened dominantly in the latter half of the century, with the most common words being central in economic vocabulary.

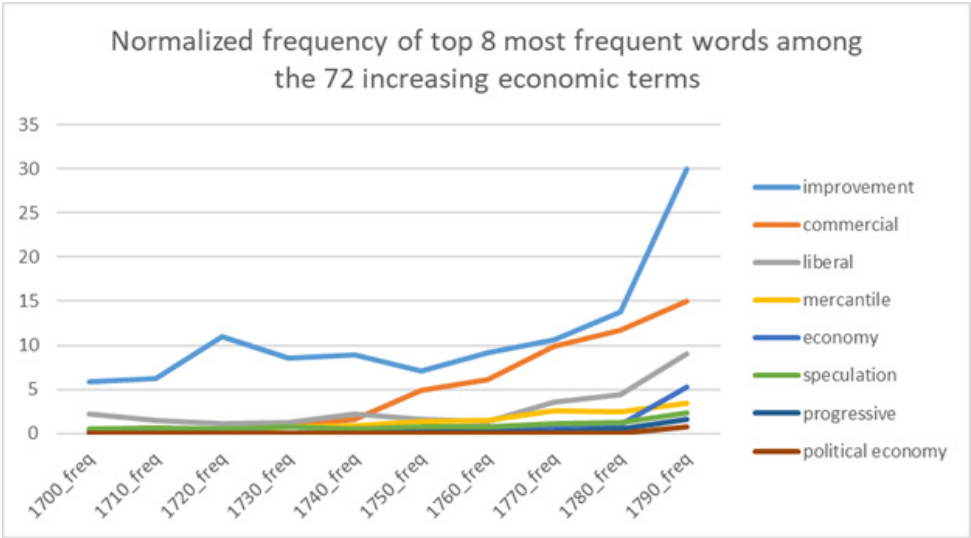


Figure 2. Normalized frequency of top 8 words among the 73 consistently increasing words identified by the PCR analysis

3.2.2 Underlying discourses

In order to identify some of the discourses which may underlie these incoming lexical items, we employ *factor analysis*, inspired by the multi-dimensional method of register analysis (e.g. Biber 1988). Factor analysis identifies “a set of latent constructs underlying the battery of measured variables” (Fabrigar 1999), in our case, a set of discourses underlying the use of certain incoming lexical items. In practical terms, we can use factor analysis to identify groups of lexical items which vary in frequency together, that is, whose frequencies in documents are correlated, so which tend to have a higher frequency in a document or a lower frequency in a document at the same time. The assumption behind the MDA methodology is that there is a functional reason why certain features tend to co-occur in the same documents. In our case, we interpret the co-occurrence of certain items of economic vocabulary to be because of their shared contexts of use, and therefore because of certain underlying discourses (*cf.* e.g. Zupardi/Berber Sardinha 2020).

The factor analysis as performed here draws heavily on the principles and solutions used by Biber (1988). Factor analysis can be performed using various factoring methods; we use the commonly-used Principal Axis Factoring. In order to make the extracted factors more readily interpretable, the resulting factor solution needs to be rotated, i.e. the axes of the factor space are changed so that separate groups of co-occurring items load more strongly on separate factors. We did this using the oblimin rotation method, which allows for a degree of correlation between the extracted factors.

Allowing potential correlations is desirable because it is not reasonable to assume that the underlying discourses are completely independent from each other. The factor analysis itself was run twice. In the first round, five factors were extracted, which was the highest number of factors that could be extracted without issue, to catch as much of the overall variation as possible. After this round, all items which did not have a loading equal to or higher than $|0.2|$ on any of the factors were removed, as such items were not very important for the description of any of the factors. Only the 28 items left over were included in the second round of the analysis. In this round, the highest number of factors that could be extracted without issue was four. All factor solutions of two to four factors were investigated. In the end, the four-factor solution was chosen because it explains the highest proportion of the overall variation and all four factors appeared readily interpretable.

Next, dimension scores were calculated for each ESTC record in our dataset. First, each lexical item was assigned to the factor(s) it had loadings over $|.30|$ on. Then, the frequencies of each of the items were standardized to their mean and standard deviation. This was done to prevent high-frequency items from drowning out the effect of low-frequency features. After this, the dimension score of each ESTC record on each of the dimensions was calculated by summing the standardized frequencies of the features associated with that dimension within that document.

In our analysis of the functions and underlying discourses of the dimensions, we make use of standard corpus-linguistic methods such as concordance lines and collocation analysis; the results of this analysis are reported in subsection 4.2 below.

4 ANALYSIS

Following the procedure described above, we extracted the four dimensions displayed in Table 1. The items on each dimension tend to be present in the same texts and absent in the same texts, but the different dimensions are free to vary independently from each other. In practice, there is a small degree of correlation between the different factors, varying between 0.08 and 0.37, depending on the pair of factors. In this section, we will first look into the individual items associated with these dimensions in subsection 4.1. Then, in subsection 4.2, we will analyze the dimensions to interpret, define, and label them.

Table 1. The four extracted dimensions and the lexical items associated with them

Dimension	Lexical items
D1	Income, expenditure, incidental
D2	Circulating medium, funded, unfunded
D3	Income, finance, financier, financial, funding
D4	Commercial, liberal, improvement, extent, produce

One thing of note is that, unlike in many multi-dimensional studies, the above dimensions only include lexical items with high positive loadings, and no items with high

negative loadings. The items with positive and negative loadings on a dimension tend to be in a complementary distribution: when the frequency of the positive items is high, the frequency of the negative items is low, and vice versa. In other words, none of these groups of lexical items is associated with a contrasting complementary group of items. The reason for this is largely technical: overall, the items included on these dimensions are relatively rare in the full dataset, which means that the most common frequency for these items in the documents is zero. Due to this, a decrease in the frequency of any set of items can not be easily correlated with an increase in the frequency of any other complementary set of items. However, the co-occurring groups of words in these four dimensions can already tell us a great deal about the kinds of discourses in which these rising lexical items were used.

In subsection 4.1, we will first explore the lexicographical metadata related to the HT ‘trade and finance’ words more generally. Then, we will focus on the incoming items included in the four extracted dimensions and explore the lexicographical metadata of these items in more detail. In subsection 4.2, we will then analyze the four dimensions, interpreting the discourses underlying the dimensions extracted in the section above using standard corpus-linguistic methods.

4.1 Lexicographical metadata

We combined the HT lemmas from the ‘society > trade and finance’ category with OED metadata. While the ID correspondences between the HD and OED are not complete, this yielded us a set of 4,084 trade and finance lemmas which could be combined with their metadata. We were thus able to examine the source languages of new trade and finance lemmas in conjunction with the dates of their first attestation. It should be noted that once a word has entered the English language from another language (e.g. *commercial*, dated ante-1687, source language Latin), its subsequent derivatives (e.g. *commercialism*, dated 1849) have English as their source language. Thus English is the source language not only of native neologisms but also of a substantial group of lemmas whose roots may ultimately be in other languages. Of the 4,084 lemmas in this dataset, with first attestations between 601–4 and 1994, 2,498 have English as their source language. The second-largest source language is French (524 lemmas in total), followed by Latin and German (290 and 59 lemmas, respectively). The number of new words per century rises to its highest point in the 1600s (813 lemmas) and actually drops in the 1700s (443 lemmas), but rises again in the 1800s (762 lemmas). French was the most prolific foreign source of new words in the fourteenth and fifteenth centuries (102 and 125 new lemmas, respectively), which is to be expected, considering the position of French in medieval Europe as “the language of the princely courts and the courts of law, of high culture (secular and religious), and of bourgeois aspiration and trade” (Putter and Busby 2010: 3). Latin reached its peak as the most common non-English source language in the seventeenth century (122 new lemmas), at a time when Latin was still a popular publishing language but vernacular publishing had already surpassed it in Britain and was starting to grow substantially (Marjanen *et al.* Under review).

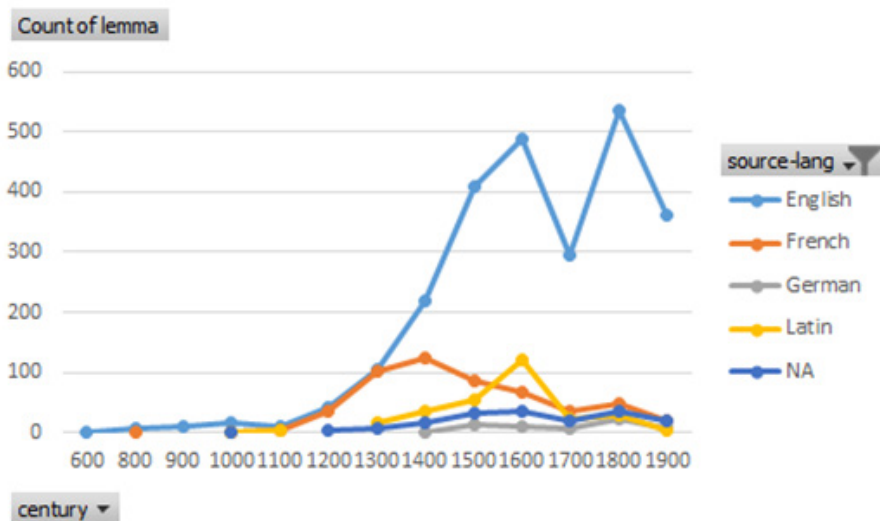


Figure 3. The source languages and first attestation dates of HT ‘trade and finance’ lemmas in languages with over 50 occurrences based on OED metadata. English lemmas include words that were derived within English from lemmas of foreign origin

We further analyzed the OED and HT metadata on the lexical items associated with each dimension. In Dimension 1, all words belong to the HT class ‘society > trade and finance > management of money’. While *income* was first attested in 1601 according to the OED, both *expenditure* and *incidental* (as in *incidental charge/expense*) are new to the latter half of the eighteenth century, first attested in 1769 and 1791, respectively. Both words are somewhat learned derivatives: *expenditure* was derived from Latin *expeditus* + *-ure*, while *incidental* comes from *incident* + *-al*, perhaps modelled on French *incidentel*. We may thus expect these items to co-occur in rather specialized registers discussing the management of money, chiefly from the 1790s onwards.

The words in Dimension 2 are found under ‘money’, ‘management of money’, and ‘financial dealings’ within the trade and finance category of the HT. While *funded* and *unfunded* were first attested in the 1760s and 1770s, respectively, the OED gives *circulating medium* a first attestation date as late as 1803, although the compound was clearly in use earlier than that based on our dataset. The noun *fund* was borrowed from Latin (*fundus*); the verb *fund* as well as the adjectives *funded* and *unfunded* were derived within English. *Circulating* likewise stems from the Latin *circulāt*, and *medium* is a direct borrowing from Latin.

Dimension 3 shares the word *income* with Dimension 1; the noun *funding*, first attested in 1735, shares the same root as *funded* and *unfunded* in Dimension 2. The rest of the words in Dimension 3 are interrelated: *finance*, *financier* and *financial* can be traced back to Old French *finance*, although *financier* is a later borrowing of the French word. *Finance* is given various categories and meanings in the HT, with first attestations

ranging from 1439 to 1866. *Financier* is first attested in the 1600s and *financial* in the late 1700s. The Dimension 3 words mostly appear in the ‘financial dealings’, ‘management of money’, and ‘fees and taxes’ subcategories of trade and finance, although *finance* has entries under ‘money’ and ‘payment’ as well.

Apart from *commercial*, the words in Dimension 4 have a considerably higher number of meanings in other HT categories than they do in trade and finance. *Commercial*, *improvement*, and *extent* were first attested in the late 1600s, 1400s, and 1300s, respectively, and their first trade and finance senses date back to the same periods. *Liberal*, however, was first attested in the late 1300s but its recorded trade and finance usage only dates back to 1816, even if the OED does give an example of the economic sense of the word developing in the 1770s; *produce* was first attested in the 1400s and had its trade and finance first attestation in 1585. As for the origins of the Dimension 4 words, *produce* was borrowed from Latin and *improvement* partly borrowed from French, partly derived within English, while *commercial*, *liberal*, and *extent* have a mix of French and Latin roots that, in the case of *liberal* and *extent*, date back to the middle ages. Overall, these words are older than the other dimension words, and have more meanings outside the economic.

It seems that the trade and finance vocabulary that increased in frequency during the eighteenth century is comparatively young and of French or Latin origin. The first attestations listed in the OED, whether for the word overall or for a particular usage, are of course not precise and may change when new sources are discovered, but they are a reliable indication of the dates by which the word or usage appears at the latest. We also looked at trade and finance lemmas whose frequency decreases or stays the same over this period and discovered that many of them, in contrast, are inherited from Germanic and date back to Old English – for example *star*, *fly*, and *sit*. However, they also tend to have several meanings that are not directly related to economic topics and somewhat rare or obscure trade and finance usage. A closer look at handpicked stable or decreasing trade and finance related words such as *traded*, *imperial*, and *pay* – the first two of multiple origins and *pay* a borrowing from French – reveals that even the decreasing ones tend to be of mixed origin, but they are typically older borrowings than the increasing dimension words above – *ounce*, for example, is partly a borrowing from Latin and dates back to Old English.

4.2 Analysis of the dimensions

A central means of understanding the dimensions is to analyze how the lexical items associated with the dimension are actually used in context. In practice, this is done using concordance lines and close reading of the texts. To start with, we searched for each of the items in three highly scoring texts per dimension from the 1790s, the period of most change in our data. Dimensions 1 and 3 share a lexical item (*income*) and also share one of the texts in this selection. Of the lexical items of Dimension 2, especially *expenditure* but also the shared item, *income*, were repeated frequently as column headings or labels in tables appended to the texts, which may explain their frequency and characterize this dimension. All three texts inspected for Dimension 1 dealt with income tax

as a means to solving state expenditure and national debt (there was only one instance of the third lexical item, *incidental*, in the three texts). The cause for said debt is also the focus of Dimension 2, in which the bills and debts that are *funded* or *unfunded* are identified as those of the navy in particular, caused by the American revolutionary war (*circulating medium* appears more rarely but is also related to the need to manage state finances). Notably Dimension 2 consists almost entirely of anonymous pamphlets; the texts surveyed here are by known authors.

The topics of Dimension 3 are much the same as 1 and 2 but with a focus on financial measures; *funding* appears repeatedly in the context of “the funding system”, also referred to as “the English System of Finance”, so this dimension seems to capture an awareness of the systemic change underway. Dimension 4 stands out as the one that deals with commerce, agricultural production, manufacture and industry, and the development of and relationships between particular (often colonial) areas.

Since it seems based on the close reading of a selection of the texts that the dimensions are meaningful and interpretable, we performed a larger-scale analysis of the dimensions using standard corpus-linguistic methods, most importantly collocation analysis and concordance lines. To do this, we created a subcorpus of all of the texts which are in the top 5 per cent on any of the dimensions. We then made use of AntConc (Anthony 2022) to analyze the collocations and concordance lines of all of the words included in the dimensions. Specifically, we searched for collocations within a window of five in both directions, and which appear at least in five separate texts and five times in total in the subcorpus.

This analysis is in line with the findings of the close reading. Dimension 1 appears to deal particularly with public income and expenditure. In the top document subcorpus, the words *income* and *expenditure* collocate with words such as *national*, *taxable*, *annual*, and *public*. *Incidental* is commonly used to refer to additional expenses or situations which may cause them, such as “other circumstances incidental to war”; indeed, *war* appears as a collocate of *incidental* in the analysis. Dimension 2 has to do with public debt. As in the close reading above, *funded* and *unfunded* both overwhelmingly relate to funded and unfunded public debt. Dimension 3 is related to the financial system as a whole, for instance, politics, committees, ministers, reports, and resources related to the financial system. Dimension 4 refers to various aspects of private commercial, agricultural, and manufacturing enterprise. *Commercial* is often used to refer to commercial actors as opposed to e.g. the landed interests, or e.g. commercial treaties; *improvement* is used most commonly in reference to the *improvement* of e.g. an *extent* of land.

To corroborate this analysis of the functions of the dimensions, we also looked at the authors and titles of the ten highest-scoring documents from each dimension. The results of this analysis also closely mirror the above analyses. The documents on Dimension 1 overwhelmingly deal with public finances and income, particularly the income tax. This dimension includes texts such as “Observations on the produce of the income tax, and on its proportion to the whole income of Great Britain” by Henry Beeke and “A review of Dr. Price’s writings, on the subject of the finances of this kingdom” by William Morgan. The texts on Dimension 2 either have to do with

public debt and credit, or the state of the nation's finances more generally. These include texts such as "An inquiry into the state of the finances of Great Britain" by Nicholas Vansittart and "Observations on the national debt, and an enquiry into its real connection with the general prosperity" by an unknown author. The titles on Dimension 3 are less focused, but many of them are still in various ways related to the system of finance and national funding, such as "A letter on the present measures of finance" by James Maitland Lauderdale or "A plan for raising the supplies during the war" by an unknown author. Dimension 4 clearly deals with trade, commerce, manufacture, and the improvement of land. These works include titles such as "A representation concerning the knowledge of commerce as a national concern; pointing out the proper means of promoting such knowledge in this kingdom" by J. Massie, and "A dissertation on the chief obstacles to the improvement of land, and introducing better methods of agriculture throughout Scotland" by an unknown author. In general, the evidence suggests the significance of Scottish influence in shaping the development of economic discourse (see Hont 2005). The majority of the authors producing the ten highest-scoring documents for each dimension are Scottish. There is a clear practical element of improvement present also in these texts that can be seen as the dominating aspect of the impact of the Scottish Enlightenment on the economic discourse in eighteenth-century Britain (*cf.* Sher 1985).

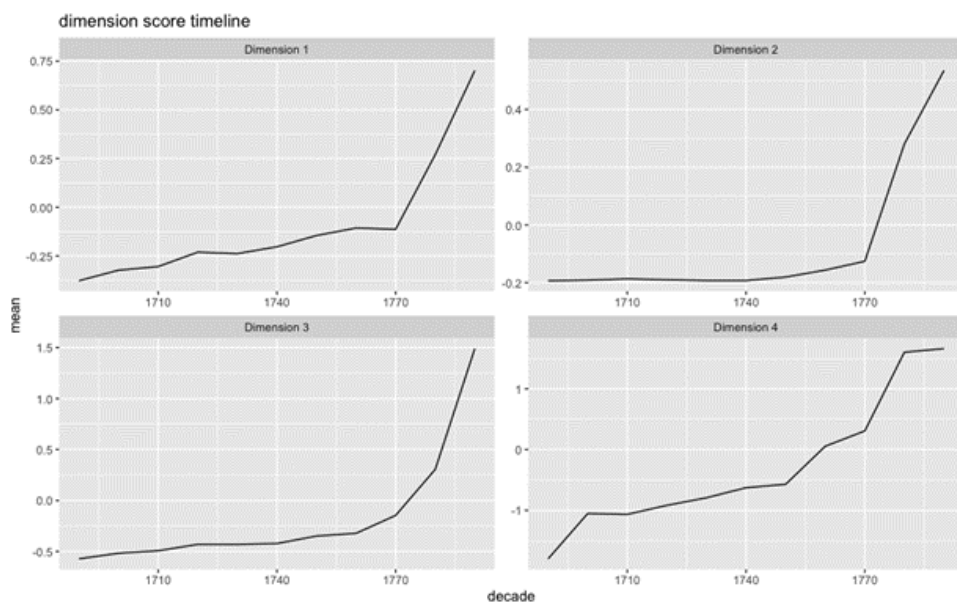


Figure 4. The average dimension scores per decade in the ECCO GKL documents for all four dimensions

However, since we are focusing on incoming lexical items, we may expect there to also be diachronic differences in their contexts of use. Since all of the items included on the dimensions are much more common in the later decades of the century, their frequent appearance may overshadow their use in earlier periods in an analysis which does not take the diachronic dimension into account. Figure 4 shows the average dimension scores of all GKL documents included in ECCO for each decade of the century. This figure shows the effects of focusing on incoming items very well, as the documents in the latter decades of the century have considerably higher average dimension scores than the documents in the earlier periods due to the overall higher frequency of the items towards the end of the century. We can also see in the figure how Dimension 4, which includes the earliest attested words of any of the dimensions, increases throughout the century; on the other hand, Dimension 1 increases slowly first but then rapidly at the end of the century, and Dimensions 2 and 3 are all but nonexistent before the end of the century (*cf.* Table 2).

In order to extend the analysis to the diachronic changes in the contexts of use of the lexical items, we created four additional subcorpora, one for every quarter-century, which included the texts which are in the top 5 per cent on any of the dimensions during that quarter-century. Table 2 summarizes our diachronic findings for each of the dimensions. Based on the analysis of these subcorpora, in the beginning of the century, generally speaking, the items tend to be used with more concrete meanings and in more concrete contexts. Towards the end of the century, their use diversifies, and the use of the items in more abstract and general senses increases alongside the more concrete ones. For instance, *income* on Dimensions 1 and 3 first often refers to the income of certain groups of people or e.g. estates or specific funds. Towards the end of the century, more references start appearing to higher-level and larger-scale issues such as national income and expenditure, including e.g. income tax, which was introduced in the UK around this time. Similarly, in the beginning of the century, *produce* on Dimension 4 often refers to what is produced in concrete terms by estates, lands, or the country. Its use also diversifies over the century, and it too is used more and more to refer to more abstract concepts such as produce of taxes, of the customs, of the state, or of labour. Of course, these patterns are only tendencies: examples of both concrete and abstract uses can be found from all time periods a word appears in. Some manifestations of these observed patterns are illustrated by Examples 1 and 2 below. Example 1 shows the word *income* referring very directly to the income of a certain set of gentry, whereas in Example 2 the words *income* and *expenditure* refer to the financial situation of the country as a whole.

- (1) There is hardly one of this new Set of Gentry, from Two Thousand Pound Fortune and upwards that do not spend near half their **Income** in foreign Wines, Linens, Silks, Laces, Tea, Coffee, and an infinite number of other Curiosities.

Anonymous (1720): *Considerations on the present state of the nation*

- (2) [H]e has consequently increased all the other property in the kingdom, if not precisely in the same proportion, certainly to a considerable amount: he has not only made the **income** of the country equal to its **expenditure**, but has also procured a surplus of a million per annum, to be employed in the reduction of the national debt.

George Pretyman (1786): *A short answer to Earl Stanhope's observations on Mr. Pitt's plan for the reduction of the national debt*

The word *commercial* on Dimension 4 also starts out as referring mostly to more concrete groups and actors, such as “trading men in the commercial world”. As the century progresses, the number of references to the country and its people as a commercial actor increases, including collocates like “a commercial nation”, “a commercial people”, or “commercial empire”, reflecting a view of the nation as a whole basing its identity on commerce. Other higher-level uses of *commercial* increase through the latter half of the century, such as “commercial intercourse”, “commercial treaties”, “commercial concerns”, or “commercial system”. These changes start taking place earlier than most of the others, as suggested by the earlier increase in the graph in Figure 4, and they point to the emergence of a more specialized and abstract discourse with regard to political economy. Economic theory addressed commercial activity as increasingly international and was less engaged in private entrepreneurship.

Table 2. A summary of the main diachronic tendencies in the contexts of use of the words associated with the four dimensions

Dimension	Beginning of the century	End of the century
D1 public income and expenditure: <i>income, expenditure, incidental</i>	Mostly private income of individuals or companies	Both private and public income and expenditure, and the relationship between the two, e.g. taxation of income
D2 public debt: <i>circulating medium, funded, unfunded</i>	Very rare	Mostly funded and unfunded debt, also e.g. funded property or funded taxes
D3 financial system: <i>income, finance, financier, financial, funding</i>	Very rare	Various aspects of the financial system and its public uses, including the funding of debts, bills, or the military
D4 private enterprise: <i>commercial, liberal, improvement, extent, produce</i>	Commerce, agriculture, manufacturing, etc. as an activity of the people	Commerce also as an international activity and a matter of national interest

In general, these changes appear to reflect the overall development of the system of economy over the century. For example, with the exception of *income*, all of the lexical items associated with Dimension 3 – *finance*, *financier*, *financial*, *funding* – were extremely rare in the first half of the century, with an increasing number of instances from the third quarter of the century onwards.

Furthermore, some of the diachronic differences represent topics which have been of particular interest or importance in specific periods. The most prominent of these is the inclusion of military matters, most importantly “navy” and “war”, in economic discourse particularly in the final quarter of the century. This vocabulary mirrors the economic competition that concretely led to war efforts in the years after the American and French revolutions.

5 DISCUSSION

This paper has focused on incoming economic vocabulary extracted from a set of words based on previous lexicographical research. These choices have naturally excluded a number of potentially interesting words whose frequency may not have consistently increased in the dataset as a whole or that may not have been identified as relevant to economic discourse in the sense of the ‘trade and finance’ section of the HT. There is therefore much to be done in future research on eighteenth-century English economic lexis. One avenue would be to identify relevant words in a more data-driven manner based on the corpus alone. Of course, the GKL itself does not represent all of eighteenth-century economic discourse, so we could also augment the corpus by identifying similar texts in the full ECCO (*cf.* Tiiponen *et al.* 2022). As mentioned in Section 3.1 above, OCR errors are an issue in the data that is yet to be fully resolved, although using the number of characters as the normalization base has mitigated its effects somewhat. Recall could be improved by using fuzzy searches, whereas improving precision could require manual consultation of the document images, as discussed with regard to the spelling variants of *economy* in Section 2.

In the previous section we identified lexical dimensions that grouped together incoming economic vocabulary co-occurring in eighteenth-century texts in the GKL. It would be of interest to relate these dimensions to different registers in the corpus: for instance, parliamentary debates could be an important register in the development of discourses of public economy (dimensions 1 and 2), and legal texts may have helped to spread specialist vocabulary to wider public consciousness. In future research, we aim to generate register information for GKL texts using both Biber’s linguistic features and machine learning methods that would identify registers through extrapolation from existing corpora with register metadata. This would also enable us to conduct a more thorough investigation into how Biber and Finegan’s (1997) finding of the divergence of popular and specialized registers from the eighteenth century onwards relates to the domain of economy and economic vocabulary.

In general, our analyses showcase how economic vocabulary grew in frequency and went through a lexical diversification. We hypothesized that this development would be related to an expansion of registers and the diffusion of economic discourse to more

popular forums as well as increased intellectual inquiry. However, our analyses only support the latter part of the hypothesis, indicating that economic terms started to appear in increasingly abstract uses and that a more specialized economic discourse emerged during the eighteenth century. The trends of a democratization and a specialization of economic discourses are not mutually exclusive, but perhaps our analysis, with its focus on incoming rather than established vocabulary, is more prone to capture the latter. The specialization is clearly related to the so-called economic bestsellers in eighteenth-century Europe (Carpenter 1975) and the consequent emergence of political economy as a field for intellectual theorizing. The obvious example of this is the lexicalization of the term *political economy* itself.

6 CONCLUSION

We investigated incoming economic lexis and the discourses underlying their rise in the eighteenth century in a collection of economic literature. We analysed the lexical metadata of economic vocabulary and found that the incoming economic vocabulary is largely Latin or French in origin, whereas the stable and outgoing economic vocabulary tends to be either of native English Germanic origin or older loans from e.g. French or Dutch, with dominant non-economic meanings. In order to identify incoming lexical items, we made use of PCR (Proportional Change Rate), a novel method for identifying periods of linguistic change and items of interest.

Using multi-dimensional analysis methods inspired by Biber (1988), we extracted four dimensions of incoming economic lexis. While similar methods have been applied to content words before (e.g. Zuppardi/Berber Sardinha 2020), we used the method for the purposes of identifying incoming economic discourses in historical textual data. We identified four lexical dimensions: public income and expenditure, public debt, financial system, and private enterprise. The development and rise of the lexis related to these dimensions can be linked to the overall development of both the economic system itself and the contemporary economic discourse in the eighteenth century. By analyzing actual usage in a large historical database of texts, our study has contributed new evidence for the hypothesis that lexical change often reflects broader sociocultural change (e.g. Allan 2015).

References

- ALLAN, Kathryn (2015) "Education in the *Historical Thesaurus of the Oxford English Dictionary*: Exploring diachronic change in a semantic field." In: Jocelyne Daems/Eline Zenner/Kris Heylen/Dirk Speelman/Hubert Cuyckens (eds), *Change of paradigms – new paradoxes: Recontextualizing language and linguistics*. Berlin: Walter de Gruyter, 81–95.
- ANTHONY, Laurence (2022) AntConc (Version 4.0.5) [Computer Software]. Tokyo: Waseda University. <https://www.laurenceanthony.net/software>
- BIBER, Douglas (1988) *Variation across Speech and Writing*. Cambridge: Cambridge University Press.

- BIBER, Douglas/Edward FINEGAN (1997) "Diachronic relations among speech-based and written registers in English." In: Terttu Nevalainen/Leena Kahlas-Tarkka (eds), *To explain the present: Studies in the changing English language in honour of Matti Rissanen*. Helsinki: Société Néophilologique, 253–275.
- BONNYMAN, Brian (2012) "Agrarian Patriotism and the Landed Interest: The Scottish 'Society for Improvers in the Knowledge of Agriculture', 1723–1746." In: Koen Stapelbroek/Jani Marjanen (eds), *The Rise of Economic Societies in the Eighteenth Century: Patriotic Reform in Europe and North America*. Basingstoke, Palgrave Macmillan, 26–51.
- CARPENTER, Kenneth E. (1975) "The economic bestsellers before 1850: A catalogue of an exhibition prepared for the History of Economics Society meeting, May 21–24, 1975, at Baker Library." *Bulletin of the Kress Library of Business and Economics, Harvard Business School* 11, 1–33.
- CROSSGROVE, William (2000) "The Vernacularization of Science, Medicine, and Technology in Late Medieval Europe: Broadening Our Perspectives." *Early Science and Medicine*, 5(1), 47–63.
- CULPEPER, Jonathan/Minna NEVALA (2012) "Sociocultural processes and the history of English." In: Terttu Nevalainen/Elizabeth Closs Traugott (eds), *The Oxford handbook of the history of English*. Oxford: Oxford University Press, 365–391.
- DURKIN, Philip (2002) Changing documentation in the third edition of the *Oxford English Dictionary*: Sixteenth-century vocabulary as a test case. In: Teresa Fanego/Bélen Méndez-Naya/Elena Seoane (eds), *Sounds, words and change: Selected papers from 11 ICEHL, Santiago de Compostela, 7–11 September 2000*. Amsterdam: John Benjamins, 65–81.
- FABRIGAR, Leandre R./Duane T. WEGENER/Robert C. MacCALLUM/Erin J. STRAHAN (1999) Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods* 4/3, 272–299.
- HILL, Mark/Simon HENGCHEN (2019) "Quantifying the impact of dirty OCR on historical text analysis: Eighteenth Century Collections Online as a case study." *Digital Scholarship in the Humanities* 34/4, 825–843.
- HONT, Istvan (2005) *Jealousy of trade. International Competition and the Nation-State in Historical Perspective*. Cambridge MA: Harvard University Press.
- JONES, Peter (2016) *Agricultural enlightenment: Knowledge, technology, and nature 1750–1840*. Oxford: Oxford University Press.
- KLEINHENZ, Christopher/Keith BUSBY (eds), *Medieval Multilingualism: The Francophone World and its Neighbours*. Turnhout: Brepols,
- LAHTI, Leo/Jani MARJANEN/Hege ROIVAINEN/Mikko TOLONEN (2019) "Bibliographic Data Science and the history of the book (c. 1500–1800)." *Cataloging & Classification Quarterly* 57/1, 5–23.
- LIVESEY, James (2012) "A Kingdom of Cosmopolitan Improvers: The Dublin Society, 1731–1798." In: Koen Stapelbroek/Jani Marjanen (eds), *The Rise of Economic Societies in the Eighteenth Century: Patriotic Reform in Europe and North America*. Basingstoke: Palgrave Macmillan, 52–72.

- McINTOSH, Carey (2020). *Semantics and cultural change in the British Enlightenment: New words and old*. Leiden: Brill.
- MAGNUSSON, Lars (1992) "Economics and the Public Interest: The Emergence of Economics as an Academic Subject during the 18th Century." *Scandinavian Journal of Economics* 94 (Supplement), 249–257.
- MARJANEN, Jani/TUULI Tahko/Leo LAHTI/Mikko TOLONEN (forthcoming) "Book Printing in Latin and Vernacular Languages in Northern Europe, 1500–1800."
- MUESSIG, Carolyn (2010) "The Vernacularization of Late Medieval Sermons: Some French and Italian Examples." In: Ch. Kleinhenz/K. Busby (eds), 267–284.
- NEVALAINEN, Terttu (1999) "Early Modern English lexis and semantics." In Roger Lass (ed.), *The Cambridge history of the English language, III: 1476–1776*. Cambridge: Cambridge University Press, 332–458.
- NEVALAINEN, Terttu/Tanja SÄILY/Turo VARTIAINEN/Aatu LIIMATTA/Jefrey LIJFFIJT (2020) "History of English as punctuated equilibria? A meta-analysis of the rate of linguistic change in Middle English." *Journal of Historical Sociolinguistics* 6/2, str. 1–40.
- POPLOW, Marcus (2010) "Economizing agricultural resources in the German economic enlightenment." In: Ursula Klein/E. C. Spary (eds), *Materials and expertise in early modern Europe*. Chicago, IL: University of Chicago Press, 261–287.
- PUTTER, Ad/K. Busby (2010) "Introduction." In: Ch. Kleinhenz/K. Busby (eds), 1–13.
- ROBERTSON, John (2005) *The Case for The Enlightenment. Scotland and Naples 1680–1760*. Oxford: Oxford University Press.
- SHER, Richard (1985) *Church and University in the Scottish Enlightenment: The Moderate Literati of Edinburgh*. Edinburgh: Edinburgh University Press.
- SHOVLIN, John (2006) *The political economy of virtue: Luxury, patriotism, and the origins of the French Revolution*. Ithaca, NY: Cornell University Press.
- TIIHONEN, Iiro/Yann RYAN/Lidia PIVOVAROVA/Aatu LIIMATTA/Tanja SÄILY/Mikko TOLONEN (2022) "Distinguishing discourses: A data-driven analysis of works and publishing networks of the Scottish Enlightenment." In: Karl Berglund/Matti La Mela/Inge Zwart (eds), *Proceedings of the 6th Digital Humanities in the Nordic and Baltic Countries Conference (DHNb 2022)*, 120–134. <https://ceur-ws.org/Vol-3232/paper09.pdf>
- TOLONEN, Mikko/Eetu MÄKELÄ/Leo LAHTI (2022) "The Anatomy of Eighteenth Century Collections Online (ECCO)." *Eighteenth-Century Studies* 56, 95–123.
- WHITTEN, David O. (1978) "Democracy returns to the library: The Goldsmiths'-Kress Library of Economic Literature." *Journal of Economic Literature* 16/3, 1004–1006.
- ZUPPARDI, Maria Carolina/Tony BERBER SARDINHA (2020) "A multi-dimensional view of collocations in academic writing." In: Ute Römer/Viviana Cortes/Eric Friginal (eds), *Advances in corpus-based research on academic writing: Effects of discipline, register, and writer expertise*. Amsterdam: John Benjamins, 333–353.

Abstract

DIMENSIONS OF INCOMING ECONOMIC VOCABULARY IN EIGHTEENTH-CENTURY BRITAIN

The eighteenth century is often connected with economic improvement. Considering the significant functional expansion of the English language during this period, driven by various socio-cultural changes, and the contemporary interest in the economy, we hypothesize that this linguistic expansion facilitated the spread of economic vocabulary to new contexts. Combining linguistic and historical expertise, we study vocabulary drawn from the 'trade and finance' section of the *Historical Thesaurus of the Oxford English Dictionary* in economic texts included in *Eighteenth Century Collections Online*. We identify incoming economic lexis based on its rate of change and apply multi-dimensional analysis to extract four lexical dimensions of economic discourse, which we interpret as (1) public income and expenditure, (2) public debt, (3) financial system, and (4) private enterprise. The lexical items associated with the dimensions are mostly Latin or French in origin, and many of them are neologisms that are first attested in the later eighteenth century, suggesting their widespread introduction into the language around that time. We show that at the beginning of the century, the use of the items that were extant then tends to be more concrete and local, with more abstract and wide-reaching contexts added towards the end of the century. This suggests a specialization of economic discourse that is related to the emergence of political economy as a field for intellectual theorizing.

Keywords: corpus linguistics, historical linguistics, multi-dimensional analysis, Eighteenth Century Collections Online, Eighteenth-century studies

Povzetek

ZNAČILNOSTI PREVZETEGA BESEDJA S PODROČJA GOSPODARSTVA V VELIKI BRITANiji 18. STOLETJA

18. stoletje pogosto povezujemo z gospodarskim napredkom. Ob upoštevanju takratnega močnega funkcijskega razmaha angleščine, ki je bil posledica raznih družbeno-kulturnih sprememb, in sočasnega zanimanja za gospodarske zadeve domnevamo, da je ta jezikovni razmah pospešil razširitev besedja s področja gospodarstva na nove kontekste. Na osnovi jezikoslovnih in zgodovinskih znanj proučujemo prisotnost besedja, ki je v delu *Historical Thesaurus of the Oxford English Dictionary* umeščeno v sekcijo "trgovina in finance", v gospodarskih besedilih, vključenih v zbirko *Eighteenth Century Collections Online*. Izposojeno besedje s področja gospodarstva prepoznavamo na osnovi njegove hitrosti spreminjanja in s pomočjo multidimenzionalne analize razločujemo štiri leksikalne vidike gospodarskega diskurza, za katere menimo, da se nanašajo na (1) javne prihodke in stroške, (2) javni dolg, (3) finančni sistem in (4) zasebna podjetja. Elementi besedja, povezani s temi vidiki, so večinoma latinskega ali francoskega

izvora, številni so novotvorjenke, ki so prvič izpričane v poznem 18. stoletju, kar kaže na to, da so takrat na široko prodirale v jezik. Pokazati želimo, da je bila na začetku stoletja raba že obstoječih leksikalnih elementov navadno bolj konkrentna in specifična, medtem ko so se v abstraknejših in širših kontekstih tovrstni elementi začeli uporabljati proti koncu stoletja. To kaže na specializacijo gospodarksega diskurza, ki je povezana z nastankom politične ekonomije kot znanstvenega področja.

Ključne besede: korpusno jezikoslovje, zgodovinsko jezikoslovje, multidimenzionalna analiza, *Eighteenth Century Collections Online*, proučevanje 18. stoletja