Review

Kira Hinderks, University of Helsinki, kira.hinderks@helsinki.fi

Writing a book that is a) accessible to 18th-century scholars unfamiliar with or sceptical of the benefits of computational methods and b) of interest to seasoned digital humanities (DH) scholars is no easy feat. John Regan, one of the key members of the Cambridge Concept Lab,[1] largely succeeds in this endeavour thanks to his openness about the strengths and limitations of his method of lexical co-association, the theoretical assumptions underpinning his work, as well as the flaws and biases of Eighteenth Century Collections Online (ECCO). Regan's linguistic analysis of important aesthetic, religious, philosophical and political concepts offers new insights into how 18th-century Britons made sense of the world that surrounded them. However, Regan is at times too modest about the significance of his findings and, in a few chapters, stops short of making an explicit intervention in the existing historiography. Despite these minor issues, Regan's book should go a long way towards convincing its readers that the 18th century is a period abounding in interesting phenomena and patterns that are only discoverable with computational methods.

The theoretical foundations of Regan's approach integrate different disciplines, most notably analytic philosophy (e.g. Wittgenstein), sociology ('collective knowledge') and linguistics (the 'distributional hypothesis' articulated by Zellig Harris and J. R. Firth in the 1950s). Another theoretical assumption, not explicitly characterised by Regan as such, is that words and concepts are functionally equal. Here, Regan

---

[1] The Concept Lab (Cambridge Centre for Digital Knowledge), see https://www.crassh.cam.ac.uk/research/projects-centres/the-concept-lab-cambridge-centre-for-digital-knowledge/.

disagrees with his Concept Lab colleague Peter de Bolla, who advocates for the existence of a conceptual 'substrate' beyond the words themselves (pp.12–13). Although the first chapter mentions the work of the Cambridge School and Koselleckian conceptual history, an explanation of how Regan's approach is aligned with or stands in contrast to these traditions does not occur until the seventh chapter and is not evident from the chapter's title. To give the reader a clearer impression of how Regan's methodological contribution is located within the existing historiography, it might have been useful to move this discussion to the first chapter.

Regan's method of lexical co-association is well aligned with the book's stated aim of 'highlighting and analysing patterns in the distribution of lexis in historical corpora' (p.1). The decision to relegate much of the method description to the appendix is very sensible and helps to make the book accessible to a wider audience. Appendix two, 'How digital tools make collective meaning visible', is undoubtedly the most technical chapter of the book: it explains how the strength of lexical co-associations is measured through a modified pointwise mutual information score (mPMI) and how lexical co-associations are more interpretable than other computational methods frequently employed in DH research, such as topic modelling and word embeddings.

The other method chapter, 'Straightening out uneven ECCO' is an excellent synthesis of the efforts of the Concept Lab – as well as the Sheffield Linguistic DNA project,[2] the Stanford Literary Lab and the Helsinki COMHIS group[3] – to identify and mitigate the effects of OCR quality issues and the database's systemic biases. This chapter ought to be mandatory reading for anyone using ECCO or similar digital databases.

Regan conducts his conceptual analyses synchronically and diachronically, investigating semantic relations over specific ten-year intervals and tracking changes across the 18th century. In the majority of chapters, Regan compares 1720–1730, 1750–1760 and 1780–1790. The selected intervals provide a good linguistic snapshot of the early, middle and late 18th century whilst avoiding some of the pitfalls of ECCO, such as the lower OCR accuracy rate for sources before 1720. Regan's arguments are reinforced by black-and-white illustrations of relevant semantic networks, as well as tables indicating a concept's 20 strongest co-associations. Regan verifies the validity of his results by comparing ECCO to other 18th-century databases, namely the Burney Collection of Newspapers and HathiTrust.

Chapter two, which focuses on the semantic networks of 'beauty' and 'beautiful', is an extended version of an article published in the *Journal of Aesthetics and Art Criticism* (2021). In this chapter, Regan shows that 'beautiful' and 'beauty' operated in two different semantic contexts throughout the 18th century. Whereas the adjective co-associates with words describing aesthetics and materiality (e.g. 'paintings', 'gardens', 'landscape', 'architecture', 'marble'), the noun is linked to words with feminine connotations

---

[2] See https://www.linguisticdna.org/ecco-ocr/.
[3] See e.g. Tolonen et al. (2022).

(e.g. 'charms', 'features', 'fade', 'bloom'). This chapter is a convincing example of how Regan's method can reveal a disconnect between everyday language and specialised academic discourses, in this case the common understanding of beauty as an element of femininity and the idea of beauty in aesthetics, as discussed by philosophers such as the Earl of Shaftesbury.

The main finding of the third chapter, '"Protestant" and the antonymic production of collective meaning', is that the concept underlying the word 'Protestant' was couched in anti-Catholicism rather than a 'substantive Protestantism' shaped around ideas of liberty, as advocated by scholars like Jerry White and Kathleen Wilson (p.49). The chief co-associations of 'Protestant' are pejorative terms for Catholics – such as 'Papist', 'Popish', 'Popery' and 'Romish' – as well as political terms that served to reinforce this hostile attitude, like 'Hanover' and 'Pretender'. In contrast, the top co-associations for the word 'Catholic' are largely related to doctrinal matters and the structure of the Catholic Church, e.g. 'heretics', 'heretical', 'Arian', and 'nuncio'.

Chapter four contains what Regan describes as a 'salutary lesson' (p.74), namely that the results of one corpus do not always reflect wider usages but are instead due to the skewed composition of that corpus. This is an important caveat that scholars should keep in mind, but in this chapter, Regan's epistemic humility leads him to downplay the validity of his findings. He notes that, in the first two decades of the 18th century, the concept of 'attention' is heavily co-associated with words describing religious practices (e.g. 'devotion', 'meditation', 'prayers') or which are religiously connotated (e.g. 'reverence', 'awe', 'hearken'). After comparing ECCO with the Burney Collection of Newspapers, Regan concludes that the link between 'attention' and religion present in ECCO does not indicate a universal understanding but is skewed by the database's large proportion of religious texts in the early decades of the 18th century. However, this proportion is a genuine reflection of the 18th-century British book market, where devotional texts remained one of the most frequently printed genres throughout the century (though its overall share did decrease over time).[4] Regan could therefore have convincingly concluded that early 18th-century 'attention' was a primarily religious notion, which shifted towards a more secular conception in the latter half of the century.

Chapter five analyses the phenomenon of 'lexical disintegration', where the co-association strength in a semantic network weakens over time. According to Regan, this trend was far less common in 18th-century Britain than the converse. The concepts under investigation are 'perception' and 'knowledge', which show an increase in frequency with a simultaneous decrease in co-associated words. The semantic network of 1780–1790 compared to 1720–1730 illustrates this striking find, but Regan shies away from giving reasons for this 'loss of semantic vibrancy' (p.85). Readers might find themselves wishing that this fascinating development had been followed up with more extensive primary source work or additional engagement with the existing scholarship on these two concepts.

---

[4] See bibliographic analyses by Suarez (2009) and Tolonen et al. (2021).

Chapters six and seven focus on the linguistic imprint of 18th-century British politics. Chapter six captures the almost formulaic 'episteme of mixed government' (p.103), which, by the end of the century, consisted of the six strongly co-associated terms 'aristocracy', 'democracy', 'monarchy', 'monarchical', 'despotism' and 'government'. This high degree of mutual co-association reflects the common practice of writers characterising the British political system as a combination of aristocracy, monarchy and democracy. This finding will perhaps not be astonishing to veteran historians of political thought, but is an important example of how computational methods can verify commonly held assumptions that were based on a small number of texts primarily written by influential authors (e.g. David Hume). The seventh chapter engages with one of the most debated ideas in the history of early-modern political thought, namely Quentin Skinner and Philip Pettit's notion of 'positive liberty' and its relationship with republican ideas of virtue (pp.119–21). Regan's analysis suggests that the idea of 'negative liberty' (i.e. freedom from interference) was more linguistically prominent in the 18th century, as 'liberty' was commonly associated with words like 'bondage', 'confinement', 'slavery', and 'tyranny' (p.122). Regan is also critical of Skinner's interchangeable use of 'freedom' and 'liberty', since the two concepts had 'different distributional profiles with regard to the prepositions which follow them' (p.140). A possible extension to this chapter could be an n-gram analysis of 'freedom' and 'liberty' to capture how idiomatic phrases reflect these different distributional profiles.

As a whole, Regan's book offers valuable new insights into the intellectual and cultural landscape of 18th-century Britain. The scholarship of the Cambridge Concept Lab at large – as well as the ongoing work of other research groups and projects around the world – illustrates the positive impact of computational methods on 18th-century studies. Continued method development and long-term funding for groups like the Concept Lab are paramount to promoting new research avenues and perspectives in the humanities.

References

Regan J. 2021. ' "Beauty" and the "Beautiful": a computational analysis of the company they kept across the eighteenth-century corpus'. In: *The Journal of Aesthetics and Art Criticism* 79:1, 88–107. https://doi.org/10.1093/jaac/kpaa008.

Suarez M. 2009. 'Towards a bibliometric analysis of the surviving record, 1701–1800'. In: Suarez M. and Turner M. (eds) *The Cambridge History of the Book in Britain*. Cambridge: Cambridge University Press, 37–65. https://doi.org/10.1017/CHOL9780521810173.003.

Tolonen M., Hill M. J., Ijaz A. Z., Vaara V. and Lahti L. 2021. 'Examining the early modern canon: *The English Short Title Catalogue* and large-scale patterns of cultural production'. In: Baird I. (ed.) *Data Visualization in Enlightenment Literature and Culture*. Cham: Palgrave Macmillan, 63–119. https://doi.org/10.1007/978-3-030-54913-8_3.

Tolonen M., Mäkelä E. and Lahti L. 2022. 'The Anatomy of Eighteenth Century Collections Online (ECCO)'. In: *Eighteenth-Century Studies* 56:1, 95–123. https://dx.doi.org/10.1353/ecs.2022.0060.