Article

ANTEDATING (IN) THE *OXFORD ENGLISH DICTIONARY*

Reporting the earliest written instance of a word is an important office of the historical lexicographer. For the *Oxford English Dictionary* (OED), documentation of earliest use has been a claim to fame, filtering into the broader culture as part of the dictionary's mythos of authority. Not surprisingly, therefore, the full-scale OED revision project undertaken since 2000 (OED3), and updated at quarterly intervals at OED.com, has cared to re-research usage evidence for every entry, supplying new instances of earliest use wherever possible. Such finds are sometimes celebrated in accompanying public statements and commentaries, such as those by OED Deputy Chief Editor Philip Durkin discussing 'a splendid antedating of "white lie'' or 'a startling antedating for *partner* meaning "spouse'''.[1]

One-upping the OED in this arena has for a long time been something of a sport for recreational and semiprofessional lexicologists—not infrequently carried out in these pages, most prolifically in the 1960s and '70s, and up to the recentest issues.[2] Today antedatings are also solicited officially under the banner of specialized 'Appeals' (a venerable OED tradition retrospectively described as 'crowdsourcing initiatives', at the moment seeking antedatings for lemmas in the alphabetical range 'M–R') on OED.com[3]; as well as a general purpose online form encouraging 'linguistic

detective work'[4]; and on social media, where the hashtag #oedantedatings has from time to time marked out proposals and discussions of earlier usages. In May of 2020 an automated Twitter account (@LOWbot) created by the present author tweeted a new potential antedating of revised OED3 entries every hour for a week, based on a much longer list of 54,000 potential antedatings drawn from simple programmatic searches in repositories of historical texts.[5]

As the scale of this (otherwise casual) enterprise implies, antedating an OED entry is often a matter of looking closely in the right places. Conversely, arriving at an antedating which will stand up to such further researches is truly a challenge. As early as the late 1970s this was apparent to Jürgen Schäfer, who predicted on the basis of meticulous statistical case studies that 40 per cent of OED entries could be antedated, though 70 per cent of these antedatings would be of less than fifty years, and therefore fairly trivial in terms of their significance to the history of the language.[6]

The present article presents new comparative analyses of OED2 (1989) and OED3 (2022),[7] which indicate that Schäfer's estimates proved reasonably accurate, at least during the first phase of OED3 revisions to OED2 entries (the Second Supplement to the OED, published between 1972 and 1986, did include many earlier citations which were subsequently integrated into OED2, but systematic or comprehensive antedating was not attempted). There may be an element of

[1] Philip Durkin, 'Shakespeare's World and updating the OED: a splendid antedating of "white lie"', *Shakespeare's World* (blog) (5 April 2017), <https://web.archive.org/web/20230201000000*/https://blog.shakespearesworld.org/2017/04/05/shakespeares-world-and-updating-the-oed-a-splendid-antedating-of-white-lie/>; and Philip Durkin, 'A Huge Find for the OED – a Startling Antedating for Partner Meaning "spouse"', *Shakespeare's World* (blog), (18 May 2018), <https://web.archive.org/web/20230605160836/https://blog.shakespearesworld.org/2018/05/08/a-huge-find-for-the-oed-a-startling-antedating-for-partner-meaning-spouse/>.

[2] Steffen Ducheyne, 'A Seventeenth-Century Antedating of the OED Entry for "*Pseudoscience, n.*"', *N&Q*, Advance Access (23 September 2023), DOI:10.1093/notesj/gjad080.

[3] <https://web.archive.org/web/20230801010907/https://pages.oup.com/ol/cus/1646166399178702002/oed-m-r-antedatings>.

[4] <https://web.archive.org/web/20231023164306/https://pages.oup.com/ol/cus/1646168222064452750/antedatings-and-other-evidence>.

[5] See D-AW, 'LOWbot goes a(n)-Antedating', *The Life of Words* (blog), (16 May 2020), <https://web.archive.org/web/20230529184914/https://thelifeofwords.uwaterloo.ca/lowbot-goes-antedating/>.

[6] Jürgen Schäfer, *Documentation in the* OED: *Shakespeare and Nashe as Test Cases* (Oxford, 1980), 65–6.

[7] All OED data is published by Oxford University Press. The analyses in this article were carried out using the XML-coded text of OED3 published in June 2022, and the pseudo-SGML-coded text of OED2 produced at the University of Waterloo in 1989. Approximately 96,000 revised OED3 entries could be referred to a corresponding OED2 entry, with another 33,500 revised entries not referable, usually because they were created as new independent OED3 lemmas from subsections of OED2 entries (e.g., for combined or derived forms, which in general were not systematically researched for earliest usage). Lemmas with OED2 first dates earlier than 1200 were not factored. While this study takes the entry as its primary lexical unit, a parallel analysis of subsequent sense antedatings within entries led to largely comparable results.

coincidence to this, as several of Schäfer's premises don't exactly apply to the general case, but in the early 2000s OED3 revised earliest dates downwards an average of 38–44 per cent of the time, with 84–88 per cent of these antedatings being of less than fifty years (Figure 1).

Importantly, the discovery of an antecedent use is not the only cause of a change in the earliest recorded usage in a revised entry. Many types of revision may lead to earlier, or indeed later, first usage dates, including bibliographical adjustments to composition or manuscript dates (such as occurred for several of Shakespeare's plays), or new manuscript editions (such as, e.g., the Hengwrt Chaucer, ed. 2003), or the merging of entries.

The incidence of revised earliest dates falling *later* in OED3 than in OED2, which I call here 'postdates' (not to be confused with the latest recorded usages), has been relatively stable over the revision project, averaging between 12 and 15 per cent. In addition to the bibliographical adjustments listed above, an entry may receive a later first date (but not an earlier one) as a result of the relabeling of an earliest quotation (e.g., as 'relevant but not directly illustrative', see ABSCESS), or its reassignment to another or a new lemma (e.g., from ACCUMB to ACCUMBING), or its suppression altogether (e.g., in ACCOMPLICE, ACCUSER), or the reorganization of an entry as a whole into separate lemmas (e.g., ABURST).

Similarly to postdates, the number of entries revised with near antedates, i.e. of less than ten years (where we would expect to find most bibliographical date adjustments), has remained stable over the project and within a similar narrow range, about 12 or 13 per cent of revised entries.[8]

Longer antedatings, however, follow an arc with two important inflections, and, as Figure 1 illustrates, lead to a congruent three-part pattern of antedating rates overall. Whereas the overall antedating rate improved gradually, if slightly, from the first updates in 2000–2007, the period between 2007 and 2014 saw a steep increase antedatings, from a trailing two-year average of 43 per cent at the end of 2006 to an average of 66 per cent by the middle of 2013. This proved to be a high-water mark for OED3 antedatings, with the average
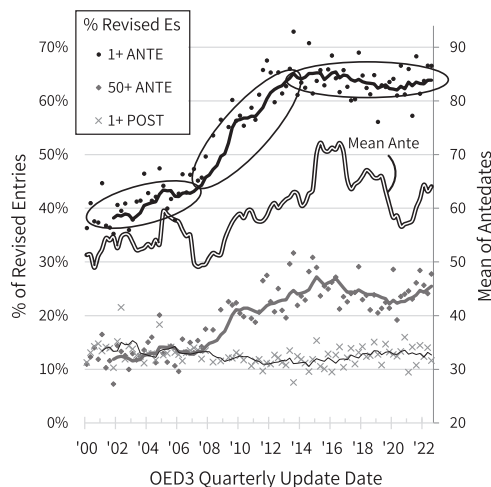

Figure 1. Revisions to OED2 Earliest Citations by OED3 Update (Rate and Mean).

remaining flat between 66 and 62 per cent in the decade since (though individual batches of updates sometimes contained substantially more, or less). This suggests that for OED3 entries revised before 2007 (about fifty-one thousand, virtually all of them located in the alphabetical range 'M–P'),[9] just under half of those with unchanged earliest dates might now be antedated, with significant numbers of antedated entries revised further.

The jump in antedatings from over 40 per cent to over 60 per cent can be explained in large part by the increasing availability of high-quality large historical text databases in this period and the lead-up to it, and to improvements in search algorithms applied to these databases, which in many cases integrated spectrum searches for likely orthographical and inflectional variants appropriate to the period covered. While OED.com does not display the particular means of acquisition for any individual quotation, the period in question saw the introduction, or significant expansion, of, to name only the most prominent: general books datasets such as Google Books (announced 2004, reaching 10

---

[8] By way of rough verification, removing from consideration those changes in earliest attribution which reference the same author reliably subtracts about half of postdates and half of near (1–9 year) antedates over the period.

[9] Because entries were revised in alphabetical order, starting with 'M', from 2000–2009, and subsequently prioritized for revision according to a number of factors, it is theoretically possible that one or more of those factors contributed directly or indirectly to the coinciding rise in antedating rate. However I have not been able to isolate any factors that correlate in this way.

million volumes in 2010); historical text datasets, especially *Early English Books Online* (EEBO-TCP Phase I 2001–2009, Phase II 2009–) and *Eighteenth Century Collections Online* (ECCO; 2005–2010); and general or regional periodical databases such as NewspaperArchive (1999–, reaching ∼900 million articles in 2008) or Australia's Trove (2009).

OED2 lemmas may be more or less susceptible to antedating due to a number of factors, including the date of publication of the original entry or its latest revision (dates for lemmas added in early OED fascicles are more likely to change than those added in the Second Supplement), the original earliest date (very old words are more likely to change than newer ones), the type of source text, and the type of vocabulary and its distribution, including its specialized or regional status.[10]

The most likely individual sources to be antedated are all nineteenth-century encyclopedias, dictionaries and lexicons, and periodicals. Most prominently among heavily cited sources, OED3 antedates three-quarters of OED2 entries first citing various editions of Webster, *Encyclopaedia Britannica*, and the *Daily News*; and more than two-thirds of those first citing Todd's *Cyclopedia of Anatomy and Physiology*, *New Sydenham Society Lexicon*, *Pall Mall Gazette*, *Westminster Gazette*, and *Athenaeum*. The most antedated sources in raw terms (apart from Webster)—all among the most quoted sources in OED2, naturally—are antedated significantly less frequently: Shakespeare forty-six per cent of the time (about average for his period), *Cursor Mundi* thirty-two, and Chaucer twenty-nine (excluding self-antedatings etc.).

Figure 2 breaks down OED3 revisions according to original OED2 first citation date, grouped by century, and the length of the change in earliest date. Among other things the graph illustrates just how unstable the dating of words to Middle English and earlier has been between these two editions, with relatively few earliest dates before 1400 surviving revision intact. Many such revisions are due to bibliographical adjustments, as the high percentage of postdates indicates. Another significant factor has been the addition of a small number of productive early sources unknown to
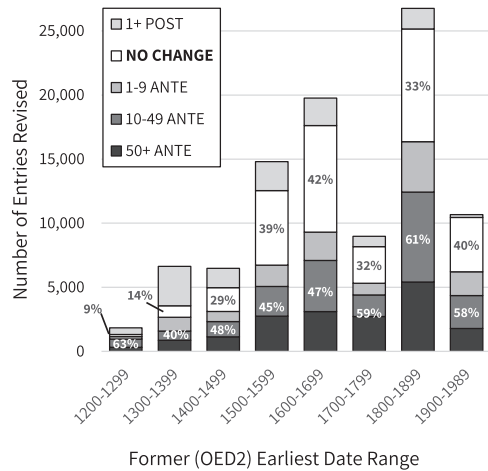

Figure 2. Revisions to OED2 Earliest Citations by Original Earliest Date.

OED2, such as Guy de Chauliac (*Grande Chirurgie*, 585 antedates), in addition to new citations from already well covered sources, such as Trevisa (323 antedates), Pecock (157), Lydgate (141), and Lanfranc's *Cirurgie* (87). Over 3,800 early first citations were added from work in the *Middle English Dictionary*, completed in 2001 and revised in 2016–2018. Important modern sources of antedates include the *Times* (450 antedates), *Philosophical Transactions* (269), the *Lancet* (223), and *Science* (171).

The percentage of entries antedated is one important measure of the revision; the length of antedating is another. Figure 1 includes a curve (keyed to the right vertical axis) showing the average (mean) length of all antedatings per quarterly update (i.e. excluding postdates and unchanged dates), smoothed to a two-year (eight-quarter) average. As with the rate of antedating, the antedating length (and thus the overall quality) improves sharply in the middle years of the revision project (though the rise is offset slightly), from a mean of fifty-one years in the initial two years, to a high of seventy-two years by the end of 2014. Subsequently this comes back down, to below sixty, before rising again to near sixty-four in the most recent two years analyzed.[11]

---

[10] I discuss differential antedatings of regional English words in detail in David-Antoine Williams, '"Alien vs. Editor": World English in the OED', *IJL*, 34.1 (Canada, 2021), 39–65.

[11] Median antedating length is offset downwards by twenty-five to thirty years, but the curve is substantively congruent.

As Schäfer pointed out long ago, very long ante-datings are of particular interest to the lexicologist, and to the historian of ideas and the social historian, among other consulters of the historical dictionary. From a linguistic point of view this is especially so where continuous usage can be documented between the previous earliest date and the new one (many long antedatings involve independent recoinages).[12] This may be illustrated with a brief selection of very long antedatings[13] of words formed with *woman* and *-ess*, as each attesting to a conceptualization operating in the culture of the time (what precisely the conceptualization may be, and what it signifies, is a matter for careful elaboration). These can be of a general kind, as may be memorialized in the literate culture, for example WOMANKIND (antedated ~200 years, to *c*1175), or WOMANLESS (~566 y, to *?c*1280); or more specifically reflective in one way or another of roles and functions in the broader contemporary culture, as BOATWOMAN (136 y, to 1707), CELLARESS (~377 y, to *a*1425), CRAFTSWOMAN (318 y, to 1568), or WORKWOMAN (~149 y, to *a*1382); or of status opportunities or lack thereof, as FREEWOMAN (~411 y, to *?c*1200), or RULERESS (~198 y, to *c*1450); and may serve additionally to evidence intersections with other conceptualizations, such as nationality or race, as recorded in ENGLISHWOMAN (~130 y, to *c*1400), SCOTSWOMAN (298 y, to 1522), FRENCHWOMAN (~255 y, to *a*1338), or CHINAWOMAN (259 y, to 1613, a decade earlier than its male-gendered homologue, and with the same added cautionary note, 'Now likely considered *offensive*').

DAVID-ANTOINE WILLIAMS (iD)
St Jerome's University, Canada

---

[12] Sometimes an earlier documentation of the grapheme per se may be indicative. A telling example is BUTCHERESS (~403y, to *a*1475), where OED2's literal nineteenth century sense is slightly antedated (OED3, 2), and a much older and now obsolete metaphorical coinage newly documented (in Lydgate's Deguileville), meaning 'A woman who procures prostitutes' (OED3, 1†).

[13] For sake of reference, overall, 9.8 per cent of OED2 lemmas analyzed were antedated by 100 years or more; 3.4 per cent by 200 or more; 1.2 per cent by 300 or more, and 0.5 per cent by 400 or more.