

landed in the hospital with pneumonia, from which I had a slow recovery. Thus my best intentions on behalf of the Society were thwarted, and *Dictionaries* 25 has appeared later than I had hoped. I thank the Society for its patience during the last few months; I know that members wait eagerly for the journal to arrive. But I especially thank contributors for their sympathy, help, and even encouragement during the last few months, which have surely been as vexing for them as for me. Some contributions that should have appeared this year will appear in the 2005 volume, and their authors have been more patient than all of the rest. Luanne von Schneidmesser has reassured members, while I have finished the volume; as usual, she has been supportive in ways that only a conscience can be.

And the 2005 volume is already in preparation — it will be in mailboxes well before the end of the year.



## The “Electronification” of the *Oxford English Dictionary*<sup>1</sup>

Charlotte Brewer

In the last twenty-odd years enormous changes have taken place in the editorial policy of the *Oxford English Dictionary* (OED)<sup>2</sup>; changes of a more profound extent and nature than ever before in its one hundred and fifty-odd year history. They have been motivated by two often coinciding factors: advances in technology that

I have not been able to find the word *electronification* in a dictionary. I use it to mean ‘the computerization [of data], so as to render it searchable and analysable by electronic means’. This article, though mostly adulatory, is in some respects critical of the revision of OED currently underway (at the time of writing, 2003, the revised text online has covered the alphabet range *M-Nipissing*). Therefore I am particularly grateful to Oxford lexicographers John Simpson (editor of the ongoing new edition), Peter Gilliver, and Judy Pearsall, as well as to the long-time friend of OED, E. G. Stanley, for their disinterested benevolence in reading and commenting on a draft; none of them necessarily agrees with any of the views expressed here. I am also most grateful to Oxford University Press and its archivist Martin Maw for generously allowing me access to the OED archives.

<sup>1</sup>In keeping with recent convention, the abbreviation “OED1” designates the dictionary published completely in 1928, whether in fascicles or bound volumes, as well as the supplement published in 1933. “SOED” indicates the four volume supplement edited by the late Robert Burchfield (1972–1986). “OED2” refers to the integrated “Second Edition” managed by John S. Simpson and E. S. C. Weiner in 1989, whereas “OED3” refers to the edition currently underway and gradually available on-line. The abbreviation “OED” then, refers to the historical project, at whatever time and in whatever form.

have transformed the possibilities of arranging large quantities of data, and a belated but thorough-going response by OED lexicographers to recent developments in lexicographical theory and practice.

### Introduction: The OED up to 1989

The history of the OED has been complex, even tangled, since its inception, but up to the last few years has been characterized by a monumental slowness and institutional inertia. As is well known, the dictionary was first adumbrated in a couple of famous papers delivered by Dean Trench to the Philological Society in London in November 1857. It was then more fully conceptualized by a number of Society members, chiefly Herbert Coleridge (grandson of the poet), in a paper of 1859, but it subsequently languished for twenty years or so, despite occasional periods or pockets of productivity, under the inspirational but over-stretched and unreliable F. J. Furnivall.<sup>3</sup> Momentum was attained under J. A. H. Murray (editor from 1879 to his death in 1915), leading to publication of its first fascicle (*a-ant*) in 1884. More fascicles were fairly steadily pumped out over the next forty-four years, when the last appeared in 1928 (*wise-wyzen*, which ended the run of *W* fascicles, *x-zzzz* having come out in 1921).

Even before this, Oxford University Press (OUP) had been preparing a supplement, heavily weighted towards the letters of the front end of the alphabet, for which the dictionary was already out-of-date — or as the Secretary of the Press, R. W. Chapman, put it, “left with a ragged edge.”<sup>4</sup> Soliciting the Vice-Chancellor’s advice on plan-

<sup>3</sup>The ‘Historical Introduction’ first printed in the 1933 edition of the OED (in part reproduced at <[http://www.oed.com/public/archive/oed2/oed2\\_hist.htm](http://www.oed.com/public/archive/oed2/oed2_hist.htm)>) describes how “as the result of a suggestion made by F. J. Furnivall to Dean Trench in May [1857],” the Council of the Philological Society appointed Herbert Coleridge, Furnivall, and Trench “as a committee to collect unregistered words in English.” Their report took the form of Trench’s two papers, which were subsequently published as a single document by the Philological Society; the second edition (Trench [1860]) can be read in the archive section of OED Online at <<http://www.oed.com/public/archive/>>. See also [Philological Society] (1859).

<sup>4</sup>Letter to Vice-Chancellor, 24 May 1933 (Oxford University Press archives, PP/1933/56. This and the other quotations from the press archives are reprinted by permission of the Secretary to the Delegates of Oxford University Press).

ning the grand lunch that was held to celebrate the publication of this final instalment of the dictionary in 1933, Chapman declared that “it is not thought practicable to provide further supplements, so that we are saying *finis coronat opus* . . . the New English Dictionary on Historical Principles does necessarily come to an end, and it may be doubted if such a comprehensive work, attempting to cover the whole vocabulary from the beginnings, can ever again be attempted.”<sup>5</sup> Chapman thought that the smaller dictionaries — the *Shorter Oxford English Dictionary* (an abridgement of OED [1933]), and Fowler’s *Concise Oxford English Dictionary* (1911) — would suffice to keep the record of the English language up to date, but the Press soon recognized that its prodigious initial investment would need more protection than this, given the industrious marshaling of words going on in the offices of its American competitor, Merriam-Webster, and elsewhere. So in the early 1950s, it began a search for someone to assemble a second supplement to record the burgeoning of massive quantities of new vocabulary since 1933 and to chronicle also the changes in usage of existing words. R. W. Burchfield was appointed to this important role in 1957 (one hundred years after Trench’s two original papers, a coincidence “cherished” by the later editor), and produced a four-volume supplement (SOED), incorporating, with some alterations, the 1933 supplement, between 1972 and 1986.<sup>6</sup>

But this second supplement did not, could not, crown the opus any more than the first. As is obvious, language develops and changes

<sup>5</sup>Letter to Vice-Chancellor, 31 May 1933 (OUP archives, PP/1933/59).

<sup>6</sup>See Burchfield (1958, 229). Not all Craigie’s and Onions’s material was included in SOED, though criteria and statistics for exclusion are nowhere stated and are not always obvious. For example, the first supplement in its first entry provides two OED antedatings for the phrase *from A to Z*, one early seventeenth-century (from a translation of *Don Quixote*), the other 1815–16 (from *Persuasion*, probably supplied by Chapman, who edited Austen’s works), neither of which is included in Burchfield’s supplement, and nor, consequently, in OED2 and the various electronic versions, all of which follow OED1 in dating the first occurrence of this phrase in 1819, in Keats’s *Otha*. The preface to OED3, the major revision of the OED currently underway, states that the “small number of brief entries found in the one-volume *Supplement to the Oxford English Dictionary* of 1933” which were omitted by Burchfield “are being reinstated in the revised text” (<http://www.oed.com/public/guide/preface.htm#general>).

at a pace with which the lexicographer finds it almost impossible to keep up. This is so even now, for online dictionaries, and very much more so for a print-based dictionary produced over a period of decades — fifty years and more for the original OED, twenty-nine for SOED. In addition, as with the original OED, the length of time it took to get the dictionary out meant that the earlier part of the alphabet was less up-to-date than the later, and there were inevitable changes and variations in editorial policy which made the eventual product uneven. A relatively trivial example of the latter is Burchfield's decision in 1973, part way through work on his supplement, to include all rather than just some jargon words — with the consequence that, alone of this group of words, *borogove*, *callay*, *callooh*, *funnious*, and *gimble* are omitted from the dictionary. More serious was the decision to include a wider range of vocabulary from non-UK sources after the publication of volume 1, with the result that words and usages from "countries such as the West Indies and even Scotland . . . have better coverage in the range H-P than . . . in A-G" (Burchfield 1975, 356).<sup>7</sup>

The prefaces to both the first and second supplements exclaim with fascination and pleasure over the developments in culture and technology that have spawned the new vocabulary crowding their pages. In 1933, the first supplement editors W. A. Craigie and C. T. Onions were particularly struck by the verbal offspring of "biochemistry, wireless telegraphy and telephony, mechanical transport, aerial locomotion, psycho-analysis, the cinema" (1933, v); while Burchfield variously remarks on buzz-words (*yumple*, *yuppie*), the "electronic environment" (SOED 4, vii and xi) (*SNOBOL*, *transputer*, *wysing* — all of which now have a quaintly dated ring), nuclear power, the sciences, wars and revolutions, and "the metalanguages of linguists and philosophers" of whose practices and language he strongly disapproved. But despite these major leaps forward in various fields of human endeavour, the lexicographical method recording the

<sup>7</sup>For a fuller treatment, see Burchfield (1974), where he also gives many inconsistencies of editorial policy within the original OED. Elsewhere he quotes Dryden's *Preface to the Fables* to explain his supplement's gradual expansion: "'Tis with a Poet, as with a Man who designs to build . . . generally speaking, he is mistaken in his Account, and reckons short of the Expence he first intended. He alters his Mind as the Work Proceeds, and will have of this or that Convenience more, of which he had not thought when he began. So has it hapned to me . . .'" (Supplement 1972–1986, 2.vii).

consequent bulges and shifts in language remained oddly the same. Burchfield records his sense of "marvel" at "the permanent value of so much of [Murray's] editorial policy, and even of his clerical procedures" (SOED 4, ix); just like their predecessors, he and his staff perused books, papers, journals, noted down usages that appeared to them novel or otherwise interesting on slips of paper, listing the author, title of work, date of edition, and a brief quotation of the word and its context, and filed them away.<sup>8</sup> Volunteers submitted copious further slips. At the same time, the lexicographers worked through the assembled slips in alphabetical order, checking the parent OED's entries, researching earlier examples of apparently new usages and words in the bowels of Bodley and other research libraries, producing concise and judicious definitions of the new vocabulary or editorial comments on semantic developments in the old. This was pretty much exactly what Murray and his staff had done all those years earlier, although they had been working without the benefit of a preceding OED.<sup>9</sup>

Because his job was to produce a supplement, an appendage to an existing and far larger work with an established authority and reputation rather than a freestanding dictionary, it was not open to Burchfield to rethink matters of on-the-page presentation and analysis any more than of lexicographical compilation. So words continued to be listed in alphabetical order, with related information presented in exactly the same form as in Murray's original (first pronunciation, then etymology, then definition plus editorial comment if any, then senses chronologically listed). Consequently there was no opportunity for consideration of other definition techniques, such as those represented in the innovative *Collins Cobuild English Language Dictionary* (1987), which broke with various sorts of traditional lexicographical convention in its attempt to supply much more contextual and grammatical information about words and their usage. Nor was Burchfield able to make any use of electronic corpora of language, although

<sup>8</sup>Burchfield (Supplement 1972–1986, 4.ix) records the procedural differences that the supplement slips were standardized in size (6 inches by 4 inches), unlike those of the parent dictionary, and that staff no longer cut up books in the process of gathering quotations.

<sup>9</sup>For a description of working methods on OED1, see the account by Onions (1928).



such use was beginning to become standard practice during the eighties, since his editorial work was by then well established — and as a conservative lexicographer and philologist he was hostile to many of the modern linguistic methods and techniques underlying the compilation of language corpora.<sup>10</sup>

Burchfield was necessarily restricted in other ways too. Ever since the publication of the OED's first fascicle in 1884, corrections and revisions of its material had poured into the OED offices in the form of letters from the public and from scholars and linguists, offprints of scholarly articles, and contributions from its own lexicographers and even publishers (Chapman and his deputy, later successor, Kenneth Sisam, were erudite and exact men of letters who took a close personal and professional interest in the Dictionary's nature and contents and supplied it with much new material themselves). Many, if not all, of these submissions were meticulously recorded and filed away in bulging folders which can still be consulted in the OED archives and the Bodleian Library. But to add to the burden on Burchfield by asking him to update and revise the material in the original dictionary, at the same time as providing a supplement of twentieth-century vocabulary, was recognized to be both unrealistic and unreasonable. When the Press began, in about 1951, its internal debate on modernizing the first edition of OED, it considered the option of revising the already published material and firmly rejected it, not least on the grounds of cost. To revise the entire work would come to around £1,000,000, while to produce a supplement of new words and senses since 1933 would cost a fraction of this amount, perhaps £30,000.<sup>11</sup> Moreover, a complete revision would take many years, while a supplement could be published within a decade or so.

As one of the two surviving lexicographers from the original OED, C. T. Onions, put it to the publishers, the dictionary had "hosts

<sup>10</sup>See *Supplement* (1972–1986, 4:x-xi). *Collins Cobuild English Language Dictionary* (1987) was the first English dictionary to be entirely corpus-based, though various earlier dictionaries had made partial use of electronic corpora, for example, the 1984 edition of the *Longman Dictionary of the English Language*, which drew on "the massive corpus of spoken and written material collected by the Survey of English Usage at University College London" (1984, xiii).

<sup>11</sup>As concluded by A. L. P. Norrington in a paper of 1954, preserved in the Publishers Papers in the OED Archives.

of wrong definitions, wrong datings, and wrong crossreferences. The problem is gigantic." He believed "it would be impossible to produce a supplement which gave an adequate treatment of all the errors in the main work . . . unless we strictly confined it to new words, senses and phrases we should get nothing done at all."<sup>12</sup> All this meant that Burchfield's supplement, although a substantial independent undertaking, left the myriad errors and imperfections of the parent dictionary — minor and excusable though they may be in comparison with the achievement of the work as a whole — virtually untouched.<sup>13</sup>

Completed in 1986, Burchfield's supplement was almost immediately followed, in 1989, by a so-called second edition of the entire dictionary. At first sight, this publication looked like a commercially driven attempt to seize a short-term publishing advantage, since it was a second edition in a limited sense only.<sup>14</sup> There was scarcely any re-editing or revision or addition of words and their treatment, but instead a re-issue in a different form of material previously available: the original OED was merged with Burchfield's supplement so as to produce an integrated, seamless text (with no indication to the reader of where OED1 left off and Burchfield began). Also added were about 5,000 new words and senses, amounting to one percent of the dictionary's total of half a million or more.

The result was not a wholly happy one for OED users. In order to keep their record of the language up to date, they had to purchase

<sup>12</sup>Oxford University Press archives: SOED/1951/14/3, and OED Revision Box file (as recorded in memo written by Dan Davin, 27 October 1953).

<sup>13</sup>In fact, Burchfield did incorporate a very small number of entries covering earlier material (such as *introspection* [1786], *intuition* [1796], novel *sb* [3b, 1639]). It is not clear why these were accepted when the majority had to be excluded; cf. his statement (*Supplement* 1972–1986, 1 xv) that "It was also decided to exclude, in the main, pre-1820 antedatings of *O. E. D.* words or senses from general English sources, since the systematic collection of such antedatings could not be undertaken at the present time." There is some evidence in the archival material relating to the SOED that the date before which new quotations might be inserted varied: one set of instructions to the staff (post-1972) specifies 1850, apparently later altered to 1830; another specifies "c. 1820 or so," but the "2" is scribbled over and "4" written over the top (OED archives: Misc/39/3v, Misc/39/4.iii).

<sup>14</sup>Compare OED1's definition, with editorial headline, s.v. *edition*, sense 3 (reproduced without change in OED2).

a "new edition," costing £1,500, of material which, in the case of many individuals and libraries, they already possessed in its original separate forms — OED1 plus Burchfield's supplement. At least one subscriber wrote to complain about this "breach of faith," since he was unable to get at the 5,000 new words without reduplicating, at enormous expense, material he had already purchased (Levit 1989, 545). Inescapably, the overwhelming proportion of entries was unchanged from the first edition (apart from the transcription of Murray's nineteenth-century pronunciation into the International Phonetic Alphabet), despite the above-mentioned flooding of the OED offices and scholarly journals with corrections and additions to the first edition. This meant that vast tracts of the dictionary were now out of date, the second edition merely reproducing quotations, etymologies and definitions which had been untouched for up to a hundred years. To the embarrassment of its lexicographers, then as at other times at the mercy of their advertising copywriters and publishing masters, the accompanying fanfare of publicity made extravagant claims for the comprehensiveness and up-to-dateness of the second edition, and few of the reviews in the popular press worked out quite how slight were its improvements or additions to the versions of OED already on library shelves.<sup>15</sup>

As it turned out, however, the second edition heralded an utterly different phase of the dictionary's history. Although in appearance and presentation OED2 seemed to epitomise old-fashioned print-based lexicography, with its twenty handsomely bound volumes, and its beautifully clear and meticulously designed page and layout, it was the product of a revolutionary technology. Murray's volumes had been electronically, rather than manually, merged with Burchfield's — an enormously laborious and time-consuming operation — and in the process all the constituent elements of the dictionary, that is to say all individual words appearing in the dictionary, in combination with the characteristics with which they were associated or the category in which they appeared (headword, part of speech, etymological, editorial, definitional, quotation author, quotation text, date), had been electronically tagged. The "project team" behind the Second

Edition was well aware of the significance of this switch from print to electronic technology, describing it as "without doubt, their chief contribution to the future of the *OED*" (1989, 1.xii).<sup>16</sup> For tagging the various elements of a word in this way enabled their retrieval according to a variety of different criteria or taxonomical principles, so as to enable both lexicographers and users to access the dictionary in radically new ways. In a word, OED was now poised to escape the tyranny of alphabetization.

In 1915, *The Times* had portrayed the historical riches buried in the OED as though they were easily accessible. "If indeed we wish to trace the history of different periods and study their innovations and ideas," the journalist wrote, "we can find these dated with curious accuracy by the appearance of the new words in which they are embodied. For just as the archaeologist, when he excavates the site of some ancient city, finds the various forms of its civilization arranged in chronological strata, so we find evidences of each past generation and its activities in the superimposed strata of our vocabulary."<sup>17</sup> This description of how we might use OED assumes, and implies, that it is a simple matter to trawl the dictionary for data on new usages, quotation provenance and dating, etymologies, and the rest; whereas in fact in a book of this size, organized according to alphabetical order of lemma, such information can only be garnered piecemeal. The lexicographers themselves probably had a good sense of how OED represented an archaeological treasure house, with hundreds of different items comprehensively described and docketed. But to the user this information was available only at the level of an individual entry, and even the intensively habitual user could not hope to construct, from an overwhelming multiplicity of individual items, the complete picture, "the various forms of . . . civilization arranged in chronological strata" that the *Times* describes. Arranging the entries not by date or content or any other intrinsically motivated taxonomy, but instead alphabetically, is

<sup>15</sup>For critical reviews of the Second Edition, see Stanley (1990) and Brewer (1993).

<sup>16</sup>The process of transferring the dictionary to a different medium, using "technically sophisticated methods, more redolent of engineering than lexicography, and unprecedented in the history of the Oxford Dictionaries," is described in the prefatory material to OED2 (1989, 1.lv).

<sup>17</sup>Quoted in *The Oxford Dictionary. A Brief Account* (1916).



distinctly unhelpful for any scholarly purpose, although peculiarly productive in other ways.<sup>18</sup>

Now, however, it is possible to think of the taxonomic principle on which you wish to arrange this vast store of lexicographical information, press a few buttons, and have answers spewed out for you — often partial, unwieldy, and at first indigestible, but susceptible to further processing and investigation. Various sorts of archaeological strata, sometimes of compelling interest and value, then loom into view, the gaps and absences in evidence often as suggestive and significant as the evidence itself.

The 1989 electrification of the OED has thus proved a remarkably worthwhile and productive venture, though initially this may have been small comfort to those users thumbing through the 21,332 printed pages of the twenty volumes and disappointed, entry after entry, by the gap between the second edition's claims, the promise implicit in that designation "second edition," and its actual achievement. But a few years later — in 1992 — they were able to disburse a further £250 and acquire a CD-Rom of OED2, subsequently re-issued in upgraded versions in 1999 and 2002.<sup>19</sup> This has propelled OED reading and consultation into the stratosphere. However, as we shall see, access

<sup>18</sup>As Roy Harris remarked of the OED some years ago, "The very convention of alphabetization simultaneously decontextualizes and recontextualizes words in a way which has no small element of surrealism in it. It makes the lexicographer 'automatically' — in the various senses of that word — a Masson or a Magritte. He becomes the agent of a poetization of the banal which is all the more stimulating for being the unsought consequence of a strait-laced professional practice" (1982, 935). W. H. Auden points to the consequences of this "poetization of the banal" when he declares that, if marooned on a desert island, he would choose to have with him "a good dictionary" in preference to "the greatest literary masterpiece imaginable, for, in relation to its readers, a dictionary is absolutely passive and may legitimately be read in an infinite number of ways" (1963, 4). Other poets have found the same: Emerson believed that "neither is a dictionary a bad book to read. There is no cant in it, no excess of explanation, and it is full of suggestion. The raw material of possible poems and histories" (quoted in OED s.v. *dictionary* 1a); and T. S. Eliot believed that "the dictionary is the most important, the most inexhaustible book to a writer" (1940, 774).

<sup>19</sup>Versions 1 and 2 of the CD-Rom are now obsolete (they are not compatible with up-to-date hardware); version 3 has been reconfigured to conform with the layout of the dictionary in OED Online. A commercial edition of the first Edition of the OED was published on CD-Rom in 1987.

to OED in its electronic form has not of itself redeemed the second edition: on the contrary, it has provided the diagnostic tools to reveal just how badly needed was the thorough-going revision of the dictionary which the second edition might have been expected to supply. This revision the lexicographers are now, finally, undertaking, in a protracted program of online publication which began in March 2000, sixty-seven years after the first edition of the dictionary was completed.

This new stage in the history of the OED has transformed the ways in which the dictionary can be both used and compiled. Its consequences can be considered under two separate (although linked) heads: first, the advantages and characteristics of the electronic versions of the OED, and second, the nature of the major revision of the dictionary now taking place. I shall take these issues in sequence below, in the next two sections. In the three final sections of the article, I examine aspects of the revisers' handling of labels and quotations (which the electrification of the OED allows us to analyze in ways which were impossible before), and I consider the implications of these features, and of OED's choice of sources, for the dictionary as a whole.

### The Electronic OED

The most obvious, and least interesting, advantage of the electronic OED is that it is far swifter to look up a single word on the computer than in one of the hard print volumes — if one discounts the time taken to switch on one's machine, access the internet or find and insert the CD, and start up the program. If one is investigating three words, or thirty, however, the benefits of speed are significantly multiplied: one touches a few keys, instead of grappling with several heavy and cumbersome volumes. The disadvantage of viewing the OED on screen is that it is impossible to compare two or more entries meticulously side-by-side, as is sometimes essential for full analysis and understanding of the lexicographers' comparative treatment of vocabulary — unless one prints out the entries, or copies and pastes them into a separate electronic document. With the CD, this option is neither easy nor transparent; OED Online, by contrast, will readily yield a clearly formatted and labeled version of its material for the printer, which one can reduce (under the "print preview" option in the "File" menu) so that each printed sheet bears 2 "pages" of dictionary matter. However,

analyzing the printouts from the online version to try to discover how and why the lexicographers have made changes to OED2 during the course of preparing OED3 is not always straightforward: even the most chalcenourous<sup>20</sup> dictionary user will blanch at the task of comparing the two different versions of the verb *make*<sup>1</sup> in OED2 Online and the New Edition Online (the former occupies 71, the latter 102 pages), and identifying the ways in which the many dozens of subdivided senses have been recast between OED2 and OED3 is a task of major proportions.

The really valuable gain in accessing the OED electronically, however, is of a different order altogether. The trouble with the printed form was that its massed ranks of alphabetically ordered items presented an intimidatingly unanalyzable front, behind whose battle lines it was logistically impossible to penetrate. How could one check the consistency, however defined, of a work treating over half a million words? How could one infer the editorial policy behind the immensely varied selection and range of quotations? But all has now changed. You can in effect eviscerate the OED, lay bare its innards and see its workings, in a way inaccessible to and probably unimaginable by its original editors.

This is not visible at the level of the individual word and its definition but instead of the supporting evidence used to illustrate the word. It is well known, for example, that Shakespeare is extensively quoted in the OED. It is now possible to confirm the laborious scholarship of Schäfer's 1980 investigations and put a figure to the number of times Shakespeare's works are cited — about 33,300.<sup>21</sup> A little more experimentation will reveal that this is greatly in excess of any other author or work. To check this requires time and persistence, but after trial and error it is possible to build up a list of more and less favoured quotation authors and sources — after Shakespeare comes Sir Walter Scott (with far fewer quotations than Shakespeare, 15,800—odd), then

<sup>20</sup>*Chalcenourous*, meaning 'having bowels of brass', is a word almost exclusively used by twentieth-century OED lexicographers of themselves; see the various bibliographic references in OED, s.v., and R. W. Burchfield (1989, 16 and note 11). The word was first recorded in Burchfield's supplement, in a definition which does not explain the lexicographical connotations.

<sup>21</sup>See Schäfer (1980). The numerous forms in which Shakespeare's name and works are cited, and the varying accompanying dates, make it difficult to be sure one has traced all quotations attributed to him.

Milton (c. 12,300), Chaucer (c. 11,700), Dryden (c. 9,000) and Dickens (c. 7,500); whereas the Bible (in various translations) is quoted just over 20,000 times.<sup>22</sup> By contrast, for example, William Blake is quoted 112 times, Christina Rossetti 133 times, Emily Brontë 68 times. The medieval period also is characterised, not surprisingly, by disproportionate concentration on the sources which happened to have been edited from manuscript by the time the dictionary was compiled, many of them in the Early English Text Society (EETS) that Furnival established for the specific purpose of feeding vocabulary into the OED. Thus *Cursor Mundi* (c. 1300), edited by Richard Morris for EETS in 1874–93, is quoted around 11,000 times, and two other EETS editions, Skeat's *Piers Plowman* and Mayhew's *Promptorium Parvulorum* (the Latin-English dictionary first published in 1499) are each quoted over 5,500 times. The OED's extensive mining of these last three works had the effect of disproportionately increasing citations for the periods in which they fall, given the comparative paucity of other sources (Brewer [2000]).

This sort of outline characterization of the OED, albeit somewhat haphazardly compiled, tells one not about the growth and development of the English language in any general sense, but instead about the literature available to, and selected by, the lexicographers. Consequently it reflects their cultural and intellectual premises. It also tells one about the reading preferences of the volunteers who provided the bulk of the quotations on which the dictionary was based, many of whom sent in material extracted from their independent reading as well as from the sources specified on the various book lists issued by the lexicographers from 1859 onwards.<sup>23</sup>

<sup>22</sup>I have calculated these figures by attempting to search for all the ways in which the authors (or below, works) have been cited in OED, using the *OED Online* search tools. Searching on the second edition of the CD-Rom produces slightly different results. See the appendix in Willinsky (1994, 209–221), for further lists and figures, which differ again, and Taylor (1993), chapter 2, for a wealth of additional comparative figures for nineteenth-century authors. Interpreting such information is not straightforward, given that some authors wrote (or published) more than others: a small oeuvre might mean less chance of being quoted in OED.

<sup>23</sup>See Knowles (2000). Burchfield several times emphasizes that "to a large extent the preparation of the final copy [of both OED1 and his own supplement] for the press was governed by the choice first made by the contributors" (1989, 84); cf. "[T]he pattern of admission was governed as much by the choice made by the readers as by any abstract principles adopted by the editors. If a reader made a slip for such an item it was likely to be included, with small regard for

There are a number of indications that Murray attempted, at various different times, to broaden the range of sources, but it was impossible for him given the immensity of his task — originally conceived as recording every word in the language<sup>24</sup> — to exert a strong control over the provenance of his sources, and balance "literary" with "non-literary," or scientific, or technical sources of a wide range of different kinds (see, for example, Murray [1977, 221–224] and Brewer [2000]). From the start, the OED had a strong literary bias (so that when drawing up the division of sources by period, the first editors chose the death of Milton as one of the division points); this bias was encouraged both by the OUP Delegates — for example Jowett in 1883 — and by dictionary users and critics, who looked to the dictionary to provide a "treasure-house" of the English language as illustrated by its greatest practitioners, especially the literary giants (about whose identity there was a consensus much better established than is possible today). The preference for literary over other types of sources inevitably conditioned the sort of dictionary that was eventually produced.

During the course of the twentieth century, and the growth of the discipline of linguistics, professional linguists, lexicographers and grammarians have — for varying reasons — fought more shy of literary sources, and sought to record and analyze "ordinary" rather than "literary" language (whatever the difficulties of defining either term, or distinguishing clearly between them). But Burchfield, in his editing of the twentieth-century supplement, always strongly resisted any disjunction between the two, and frequently stressed the importance of including "great writers" in the OED. For this devotion to literary sources he was taken to task, by both internal and external critics, but according to his own account at any rate he seems to have held firm. Thus he deflected OUP's internal objections to his inclusion of eccentric poetic vocabulary and usage in his 1962 sample material for the supplement (e.g., T. S. Eliot's *loam-fee!*), resisted his staff's alleged loathing of poetry ("my staff (I don't know about anyone else's) have a genuine

consistency in comparable words, or in words drawn from other writers, in other parts of the Dictionary. Conversely a word that was not copied by a reader had little chance of inclusion since the editorial staff would almost certainly be unaware of its existence" (1989, 89).

<sup>24</sup>"The first requirement of every lexicon is, that it should contain *every word in the literature of the language it professes to illustrate* [original italics]" ([Philological Society] 1867, 2).

horror of poets. I love poetry and poetical use has been poured into the Supplement, because it is my own preference compared with that of my colleagues") and stated very firmly in the last volume of his supplement that the failure of descriptive scholars "to quote from the language of even our greatest modern writers, leave[s] one looking at a language with one's eyes partly blindfolded."<sup>25</sup>

But just how literary is or was the OED, and how significant is the literary bias? How has the dependence of the dictionary on voluntary readers, possibly with eccentric interests, affected the way in which it represents the English lexicon?<sup>26</sup> In order to assess this, we need to put together a picture of some sort of what Murray's sources were, just as we need to do this for Burchfield's sources and for those of the present-day lexicographers. To put it another way, the OED is only as good as the sources from which it is compiled. This is one of the three primary criteria by which it may be assessed and judged, the other two being the thoroughness and accuracy with which sources are read (whether by voluntary readers or the lexicographers themselves), and the use made by the lexicographers of the evidence derived from the sources.

Hitherto, it has not been possible to establish hard information on any of these three different lexicographical factors. But there are now many ways in which one can cut interrogative swathes through the data which the electronic versions of the OED render accessible for the first time. For example, one can ask how many quotations the OED records for each of the decades (or years) that it covers, from 1150 onwards — a laborious but perfectly feasible operation, whose results can then be modelled as a graph.<sup>27</sup> Such a graph turns out to have

<sup>25</sup>For *loam-fee!*, see Burchfield (1989, 11–13); for his staff's fear of poetry, see Burchfield's remarks in the transcription of a discussion following his delivery of a paper on "Aspects of short-term historical lexicography" (Pijnenburg and de Tollenaere 1980, 271–279 and 280–286); and for objections to the techniques, practices, and vocabulary of "great marauding bands" of present-day linguistic scholars, see SOED (1972–1986, 4.xi) and Burchfield (1989, 10–11).

<sup>26</sup>Some of the most prolific contributors chose to send in slips related to particular fields. For example, Marghanita Laski, who submitted as many as 250,000 slips, read detective novels intensively, while Vincent St. Trounbridge sent in hundreds of terms relating to drama. It is possible that SOED's intensive coverage of the vocabulary of surfing reflects the special interests of an individual reader (Baker 1988, 148–153).

<sup>27</sup>See Brewer (2000, 48–49 and 57–58). Schäfer (1989, 43–54) was the first person to construct such a graph (covering the years 1475–1850), although he



notable peaks (for example, in the last decade of the sixteenth century) and troughs (for example, over much of the eighteenth century). It is tempting to think that these represent the varying rate of word coinage over the centuries — but they don't, of course: or at any rate, they needn't. Instead, as with the representation of individual authors, this data primarily registers what the lexicographers chose from the material available to them to put into the dictionary. Not surprisingly, there is often a correlation between peaks in word recording, and the particularly intensive excerpting of an individual source or sources.<sup>28</sup>

This reinforces the point made above. The OED reflects the sources chosen by its lexicographers and readers. Both nineteenth- and twentieth-century lexicographers were limited by constraints of time and resources, and had to be dependent on material not always accurately or thoroughly assembled. They were not, for one reason or another, systematic in their choice or use of sources. They were inevitably reliant on a lexicographical method which traced words and their usages independently of each other, and which completely ignored synchronic relationships, despite the fact that those relationships may have decisively influenced the semantic developments which their diachronic method records.

The electronic forms of the OED enable us to discover, investigate, and attempt to understand all these things in ways which, as suggested above, were probably unimaginable by the previous

took his data from the OED bibliography published in 1933, not from the text of OED itself. Remarkably, the results now available confirm his preliminary findings.

<sup>28</sup>For example, there are (according to the online search facilities) 9,273 quotations in OED2 for the decade 1511–20, compared with more than double that amount, 21,086 quotations, for 1521–1530. This steep rise of 11,813 quotations is nearly half accounted for by a single text, Palsgrave's *Leclairissement de la langue françoise* of 1530, from which the lexicographers took 5427 quotations (according to the second edition of the CD-ROM; the identical search on OED Online produces a different total, 4621 quotations — still substantial). Schäfer attributes the paucity of eighteenth-century evidence in the OED to the loss of the slips for this period, which had been prepared largely by American readers; but Murray made it clear in 1897 that in fact what had happened was that "The American scholars promised to get the eighteenth-century literature taken up in the States, a promise which they appear not to have to any extent fulfilled" (Schäfer 1980, 53 and Murray 1880–1, 123–124).

editors.<sup>29</sup> For the lexicographers themselves, however, the primary importance of the electronification of the dictionary is that it enables diagnosis of what happened in the past, and accurate measurement and analysis of what they are doing in the present to revise the original OED and bring its record up to date.

When OED2 was published, it occasionally seemed that the editors themselves had underestimated the monumental task of revision necessary. "It is a matter of common knowledge that many elements of the original OED require revision," the editors commented at the start of their introduction (1989, iv–lv). But they felt that "This new edition represents the first, and almost certainly the most arduous, step towards [the] goal" of full revision and updating. As it has turned out, electronification has pointed the way to many more, and more arduous steps, than they may originally have supposed. The new versions of the OED make clear the extraordinary variability of quoted authors, sources, and periods, indicating that the job of full revision and updating is an enormous one.

### OED Online and the New (Third) Edition

Work on the third edition of the OED was in hand before the second edition was published. A brief description of the project, together with a public appeal for help, was announced by the editor John Simpson in a letter to the *Times Literary Supplement* of November 15<sup>th</sup>, 1993. Completion, originally planned for 2010, has since been deferred; Simpson has recently suggested that "we expect to complete the main cycle of revision in twenty years or so, depending on a number of factors (budget, growing experience of staff, new computer routines, etc.)"<sup>30</sup> Meanwhile, OED Online was launched in March 2000 as an Internet site, giving subscribers access both to the electronic version of OED2 and to the revised portions of the dictionary as they are successively completed. By October 2002, the lexicographers had recast

<sup>29</sup>The new possibilities for research opened up by the electronification of OED are still to a large extent under-investigated. A notable exception is Taylor (1993); see also Fowler (1998, 333–350).

<sup>30</sup>Communication to the author, June 2003.



words and entries in the range **M-monomisher**, together with sundry items elsewhere in the alphabet, amounting to perhaps 5% or so of their eventual total; by September 2003 they had got as far as **Nipissing**.

The first thing for a reviewer to say is that this new edition does, at long last, represent a major leap forward in lexicographical practice and representation, unparalleled in the OED's history to date. The site is exceptionally well-designed and in many respects, not least visually, a great improvement on the first two editions of the CD-Rom. It is intuitively easy to use, and offers a substantial range of information and resources, from extensive archival material on the history of the dictionary to sophisticated search options. Switching between one screen and the next is swift owing to the limited graphic content, and a site map (at <http://www.oed.com/public/contents.htm>) clearly presents an outline of the information available. The main elements of Murray's page and entry lay-out are by and large preserved, in traditional lexicographical form, but one can choose to turn these features on or off — pronunciation, spellings, etymology and quotations — so as to clear the screen. An additional optional feature, extraordinarily useful, is a datechart which represents the distribution of quotations in graphic form, so that large gaps (or not) in documentation spring immediately to the eye (although such visible variations invite questions that are not easy to answer. What does one infer from a gap in documentation of 50, or 100, or 200 years beside a cluster of quotations from neighbouring decades? That a word was used more often at some periods than at others, or that the lexicographers have not found an example? (See further below).<sup>31</sup> The site designers have thought carefully about how to exploit the new possibilities provided by the electronic medium for viewing, accessing, and cross-referencing information, and there are many imaginative and helpful options such as the ability to view contiguous lemmas listed by date or by entry. The help menu includes a tour (free to non-subscribers) explaining quickly and accessibly how to use its main features. The site is constantly updated, so that even a frequent user can be surprised by the new material appearing on the screen.

A particularly valuable feature of this new edition and its ongoing revisions is its historical transparency: most changes to the site are

recorded as they are made. These are listed in two areas on the OED Online "help" facility. "Quarterly Updates to *OED Online*" charts the successive releases of batches of revised and new vocabulary (for example, **miivy-momisher** was released in September 2002, **mis-mitzvah** in June 2002, **mid-Mirzapur** in March 2002, and so on), and also itemizes and links to such things as the brochure and worksheet provided in December 2001 and the advanced search tools added in January 2002. (These new tools have powerfully increased the complexity and hence the value of the questions one can now ask of the OED in its online form, although they still, unfortunately, do not match the facility on the CD-Rom which allows one to scroll through lists of the variant forms in which quotation authors and quotation works have been cited in the dictionary).<sup>32</sup> It is also possible to access lists of the entirely new words that have been added to the online version, together with the date of their inclusion (for example, **mixability**, **moaner's bench**, **mobile vulgus**, **modern**, **body-swerve**, **clunky**, **dysfunctional**, **dyspraxic**, **lurve**, and **toe-curling**), both those across the whole alphabet range and those within the specified ranges under systematic revision, as well as the entries to which new meanings have been added (for example, **abductee**, **black water**, **centaur**, **cringe**, **exfoliate**, and **spacer**).

The importance of this transparency cannot be overrated, since it allows the dictionary-user to identify which changes and additions have taken place at which stage in the process of compiling the revisions to the dictionary. If you take the view that the OED is not an

<sup>31</sup>The date chart does not appear, unfortunately, on the printed-out form of an entry.

<sup>32</sup>For example, Carlyle's work on the French Revolution is cited in (at least) five different ways: *French Rev.* (3 quotations), *French Revol.* (1 quotation), *French Revolution* (1 quotation), *Fr. Rev* (1 quotation), *Fr. Rev.* (1491 quotations), *Fr. Revol* (7 quotations). His biography of Frederick the Great receives similar treatment: *Fred. Gt.* (9 quotations, dated between 1858 and 1862), *Fredk.* (1 quotation, dated 1865), *Fredk. Gr.* (4 quotations, from 1864 and 1865), *Fredk. Gt.* (1244 quotations, apparently from 1858 to 1865), *Fredk. The Gt.* (1 quotation, dated 1865). (These figures are taken from the second version of the CD-Rom.) There is a wild card facility on both CD-Rom and the online version, which will for example turn up *Rev*, *Revol*, and *Revolution* if you type in *Rev\** as your search term. But it is a major loss not to be able to see in advance what the variations in citation have been, and consequently to be unsure whether or not you have been able to guess what all of them are or may be. The same applies to all the variant forms by which an author's name has been listed, which can be scrolled through as a list on the CD-Rom but are not accessible on OED Online.



impartial record of language and its users, but instead a selection governed by all sorts of individual factors relating to the lexicographers concerned and the conditions under which they worked — including their own views on and assumptions about language and what their job of dictionary-compiling entailed — then you want to have as much information as possible about who did what to the dictionary, and when. This is something OED Online implicitly recognizes by documenting its own actions in this way, and also by clarifying the bewilderingly various stages through which it has passed since its initial electrification in 1989 — six different forms since 1986, representing the progression from print to online medium, via various stages of addition and revision.<sup>33</sup> Helpful as this is, however, the differences between these stages (up to OED3, that is, which completely recasts all entries) are nothing like as substantial as those between the first edition and Burchfield's supplement, the undifferentiated merging of which was one of the regrettable features of OED2. This means that anyone interested in the first edition's treatment of vocabulary, as compared with Burchfield's, continues to need the printed versions of both OED1 (including the 1933 supplement) and Burchfield's supplement to hand, so that one can check whether apparently first edition material, whether in the definitions or quotations, is truly that, or was added by Burchfield — and also distinguish between Burchfield's additions and those of the OED2 compilers.

So how do the new entries compare with the old? Examining the various different versions side by side — or rather, successively, since simultaneous viewing of the various chronological states of the

<sup>33</sup>See <<http://www.oed.com/public/guide/citing.htm>>. The page does not mention the three CD editions, although it is important to distinguish between each of these, and between the CD Rom's and the online version: identical searches in the different electronic media often produce different results — attributable, presumably, to various sorts of technical explanations (see, for example, n28, above). OED3's transparency is not complete. The New Edition entries are in awkwardly draft form, and many of them have been altered since they were first released (for example, to change labels, quotations, ordering, and wording of definitions, etc.). The preceding versions, which one may have cited as from a stable and consultable authority, disappear without trace. Thus I possess a print-out of a revised entry for the verb *make* dated June 2000, which has been irrecoverably replaced by the current online entry, dated September 2002. The online edition is a continuously evolving organism, and the editors are of course right to take advantage of the means to correct and revise as they go. The obliteration of its successive stages is however disconcerting to those used to the stability and permanence of print.

OED is not practicable on a standard-sized screen — indicates that the third edition of has honorably recognized the need to deliver the root and branch reworking of the first edition which was eschewed by the second edition. It is clear that a wholesale revision has taken place. The semantic structure of each entry has been reconsidered and in many cases recast, so that identification of the various senses of a word may be partially or completely different. In all necessary cases, which in practice means almost all cases, surviving definitions have been rewritten in contemporary English, replacing the late nineteenth- or early twentieth-century locations that now look quaint, out-dated, and/or, for one reason or another, unsuitable.<sup>34</sup>

The recasting of entries seems often to be driven by the accumulation of additional quotations, to which a vast amount of well-directed energy has been devoted. Every student of pre-contemporary texts will have come across examples of words which antedate the OED's quotation evidence, and many users have provided the lexicographers with evidence and lists from the publication of the first fascicles onwards. For the first time, this evidence has been drawn on and incorporated into a new version of the OED, and also, evidently, backed up with independent research.<sup>35</sup> As a result of this process, the

<sup>34</sup>A random example is the rewriting of the definition for *magnanimous*, a verb supported with a single seventeenth-century quotation: "To render high-souled; to cheer, inspire" has been replaced with "To cheer, inspire, give courage to (a person)." Care has been taken to eradicate sexist definitions: compare for example the respective entries for *master* in OED2 and OED3. There may also have been some systematic pruning of sexist quotations: Burchfield's choice of "the Dry Martini...is a drink that will quickly separate the men from the boys and the girls from their principles," a quotation from a 1968 copy of *House and Garden* which he used to illustrate the phrase *separate (or sort out) the men from the boys* has been excised in OED3. This is an area which would repay further study: have all undesirable quotations been removed? Even if they provide important evidence of usage at a particular date? Is it desirable to rewrite the dictionary in this way, given that the lexicographers' choice of quotations gives us valuable information about their criteria for inclusion and exclusion, and also about the culture of their day?

<sup>35</sup>To verify this, I made preliminary investigations only, since a full study would be a substantial undertaking. I noticed many places where the revisers have added OED ante- and post-datings published in, for example, *Notes & Queries*, but only a few examples of omission. These I sent to the editors, who immediately incorporated them in the *Online* edition — a notable instance both of the advantages of online lexicography and of the receptiveness and efficiency of the current lexicographical team.



body of quotations has been reconfigured, with some alterations (both additions and subtractions) so as to illustrate more tellingly the words and senses exemplified, and also to even out (or so it often seems — see further below) the chronological spread of attestation. Existing quotations have all it appears been checked, and in some cases redated. In general, there has been a major drive to tidy up and standardise bibliographical referencing — no insignificant task.<sup>36</sup> Other categories of treatment — for example, on spelling forms, etymology, phonology, word labeling (see further below), and the occasional explanatory headline — have all been overhauled and in many cases expanded, changed, or rewritten. The result is that this version of the OED has left clear water behind it. Readers wanting to get a representative idea of the scope and detail of the sweeping changes that have been introduced may wish to examine old and new versions of, for example, *magic*, *martian*, *martyr* — or indeed *make*, possibly the most substantially rewritten and reworked item to date (although the changes made to this verb are not susceptible to casual analysis, since the entry is too massive and too detailed to be apprehended on screen).<sup>37</sup>

Perhaps most pressingly of all, OED3 is engaged in bringing the record up to date. The huge number of entries for words still current, but for which the latest quoted evidence was still a hundred years or more out of date, has now — over the revised range that is — virtu-

<sup>36</sup>As every user will know, the first (and hence second) edition of the OED, appearing as it did over many years, and under often difficult conditions and successive hands, could not maintain consistent bibliographical standards. Thus many works were allotted different dates on the different occasions on which they were cited, and authors and works were often referred to by different titles and/or abbreviations. Burchfield was more consistent than his predecessors, but by no means perfectly so (see, for example, Brewer [1993, 329n5]). Simpson discusses bibliographical regularization of OED citations at <[http://www.oed.com/public/guide/preface\\_6.htm#bb](http://www.oed.com/public/guide/preface_6.htm#bb)>.

<sup>37</sup>The OED3 version of *martian* is evidently superior to that of its predecessors, but should the revisers have recorded its use as a poetic term, current in the wake of the publication of Craig Raine's *A Martian Sends a Postcard Home* (1979)? Good quality print examples of its usage can be found in a range of standard works on poetic diction and technique (for example, Morrison and Motion [1982]), and the OED has in previous versions made a point of recording literary critical terminology, especially in relation to the writer with whom a term is associated. See Burchfield (1989, 70, on T. S. Eliot) and his supplement's treatment (reproduced in OED2) of *negative capability* (s.v. *negative* 8c), *ambiguity* (s.v. 3b), *practical criticism* (s.v. *practical* 6), and the like.

ally disappeared. This anomaly had been one of the most disfiguring aspects of the second edition, belying its claims to be "authoritative," "up to date," and "comprehensive," since for its twentieth-century component it had relied almost exclusively on Burchfield's evidence as printed in SOED.

But Burchfield's brief, as we have seen, had been only to identify and record new words (or new senses of existing words) omitted by the first edition of OED, or emerging only after its publication, and not to update the quotation record of words for which there was at least nineteenth-century evidence — for as he himself pointed out, "for the earlier letters of the alphabet such a policy would have entailed the addition of late-nineteenth-century or of twentieth-century examples for virtually every word and sense listed in the Dictionary." Consequently, OED2's attestation and treatment of pre-existing words — i.e., the bulk of the English lexicon — during the course of the twentieth century was extraordinarily thin, and its reprinting of out-of-date definitions sometimes disconcerting or even absurd. (Perhaps the most egregious example I have come across is the definition of the Conservative Party as 'one of the two great English political parties' — the other being, evidently, the Liberal Party. First published in 1891, this statement has been faithfully reproduced in every edition of the OED since, up to and including OED2).<sup>38</sup> This major defect in the OED has now begun to be remedied. Entry after entry in the portion of the alphabet range so far revised displays new quotations from twentieth- and indeed twenty-first-century sources, repairing the gaps left in documentation for the last 100-odd years. Taking the range *M-mon-nisher* as a sample, and using OED Online's search mechanism, it is possible to calculate that, whereas Burchfield supplied 19,563 quotations for this section of the alphabet, the new revisers have supplied

<sup>38</sup>Burchfield's treatment of political parties (and consequently that of OED2) is not straightforward. There is no entry for *Conservative Government* or *Conservative Party*, for example, although the former term occurs 12 times, and the latter 61 times, in the second OED (in definitional text s.v. *conservative* (n.)) 2a but elsewhere usually in quotations), and the party is treated appropriately by the first OED s.v. *conservative* (adj.) 2a. *Labour Party* was omitted from OED1 (the *Leisurely* fascicle appeared in 1902), but was treated by Craigie and Onions in the 1933 supplement, who placed the term and its definition in the ragbag category of attributive uses of *labour* and provided 6 quotations dated between 1886 and 1922. Burchfield reproduced their definition, added four more quotations, and kept the term in the same minor position (sandwiched between *labour-pains* and *labour relations*).



37,639. In other words, they have nearly doubled the number of quotations from recent sources.<sup>39</sup> It is difficult to overstate the value of this material (or the extent to which it was overdue).

Burchfield went on to say, "Our policy depends upon the realization by users of the Dictionary that any word or sense not marked "obs." or "arch." is still part of the current language." This advice, however, was unsatisfactory. One has only to turn over a few pages of OED2 to find example after example of words quite unfamiliar to a mid- to late-twentieth-century user, not marked "obs." or "arch." but quite clearly NOT "part of the current language" of the time. Most of these (so it appears) are now being caught by the third revision. The lexicographers have presumably looked for and failed to find subsequent quotations for numerous such words documented in OED1 with nineteenth-century quotations but left untouched by Burchfield in the 1970s and 1980s, and consequently also by OED2 in 1989 (for example, *magrinic*, *magrinst*, *magism*, *maginase*, *magueliferous*, *maidenism*, *manifactive*, and countless others), and at long last applied the label "obsolete" (in some instances, especially where Murray or his co-editors marked the word "rare," it may well have been obsolete for a hundred years or more). The same has been done for those words (far fewer in number) documented in OED1 with only pre-nineteenth century quotations (sometimes only one), not then identified as obsolete, and also left unmarked by Burchfield (for example, *mickening* a, last quotation in OED1 [as also in OED2 and OED3], 1608; and *melonist*, 2 quotations in OED1 [OED2 and OED3, too], dated 1629 and 1727).

Conversely, some words labelled archaic in the first edition, and again left untouched by Burchfield, have had the label removed — thus *misdoubt* (the noun, not the verb), is supplied with three twentieth-century quotations, and *misenter* now has its single OED1 quota-

<sup>39</sup>The M fascicles for OED1 were published in 1904–1908, and Burchfield's supplement volume covering words beginning with M was published in 1976, so any OED2 quotations for words over the range **M-nomnisher** between 1908 and 1976 must have been inserted by Burchfield (barring a few possibly added by the OED2 compilers — I have disregarded these as numerically insignificant). I searched for the date range 1909–1976 in "quotation date" on OED Online, and counted the number of quotations for the range **M-nomnisher**, to get the number of quotations inserted by Burchfield. I then clicked on the button giving the corresponding quotations for the new edition, and again counted up the number occurring within that alphabet range.

tion, dated 1675, sandwiched between one of 1598, and one of 1999 (leaving an odd gap).<sup>40</sup> And some words which, one might have thought, should have been labelled obsolete or rare in the second edition have now been shown to have had a new lease of life. Thus *magnanerie* 'silk-worm house' had two quotations (1887 and 1885) in OED1, was passed over without comment by Burchfield, and now in the third edition is demonstrated as having been both earlier and later used, with additional quotations from 1835, 1966, and 1969.<sup>41</sup>

The wealth and variety of differences and revisions between OED3 and its predecessors make it difficult to form a clear idea of the character of the new OED as it unfolds before us. One way of grappling with this problem is to examine a particular area of revision, however limited, to see what hints it can give us about the dictionary as a whole. In sections 4 and 5, below, I look first at editorial labels, and then at quotation numbers and distribution, to try to form a view of the aims and qualities of the OED lexicographers' massive new undertaking.

### Editorial Labels

It is well known that there is a wide range of editorial labels in the OED and that many of these are problematic. Inevitably, over so long a period of compilation, labeling practices changed and

<sup>40</sup>Surprising in view of the recent rebirth of the verb *enter* (not yet treated by the revisers) in relation to electronic data — a sense recognized in the rewriting of the definition for *misenter* to read "To enter erroneously, esp. in a book, register, *database* [my italics], or other record." *Misenter* is defined by Merriam-Webster's dictionary of 1913, indicating continued usage in the nineteenth and twentieth centuries.

<sup>41</sup>The word is unlabeled, but that the 1966 quotation is from a volume of poetry by Kenneth White, and the 1969 one from Nabokov's *Ada*, may give one pause for thought: are these isolated examples of literary resuscitation, possibly from the pages of the OED itself, a phenomenon occasionally identified by Burchfield (cf. his remarks on Auden's revival of *balloping*, or Joyce's of *peccaminous* and *cessile*, reproduced in OED2)? It would be helpful to have some editorial comment on this possibility, as with many other unlabeled words whose use is sparsely illustrated, from unusual or eccentric sources, over the last 100 years. John Simpson points out to me that this is supplied in the case of 6 words (*malalent*, *melpomenish*, *menalty*, *muskin*, *nan*, *nanmick*) all identified as (revived) "from dictionary record" — two of them by Auden and two by Joyce.



developed. Consistency, however desirable, was never a practicable possibility.<sup>42</sup> In his introduction to SOED, Burchfield tells us that the "system of labelling is unchanged" between OED1 and his supplement, and that "it would have been inappropriate to have a different system in the Supplement from that used in the Dictionary itself" (1972–1986, 1.xvi). This statement glides over the inconsistencies of labeling in both OED1 and SOED, and also obscures the fact that Burchfield did in fact treat labeling, of various kinds, differently from his predecessors — most strikingly in his introduction of Johnsonian *ipse dixit* from volume 3 onwards.<sup>43</sup> The Third Edition revisers have thus had their work cut out to regularize and systematize editorial labels, an especially important element in an electronic dictionary where users can be expected to employ label tags, and status descriptions of one sort or another ("vulgar," "regrettable," "loose," "poetic"), as one of the ways in which to organize searches for various different sorts of material. For example, how many of the quotations from Auden are of words judged to be "historical" or "archaic"? What words occurring in a particular play of Shakespeare's are still in use in the twentieth century, but have shifted in register in some way or other? Editorial labeling sheds all sorts of interesting light both on words and their connotations and also on the attitude, sometimes unconsciously expressed, of the lexicographers towards their material.

The Third Edition revisers' policy has evolved gradually, as the present writer has found to her cost. When I first investigated OED3's treatment of a number of labels ("historical," "archaic," "obsolete," "rare," and others) over a sample stretch of text, I found a number of troubling inconsistencies which I wrote up in a draft version of this article. In essence, I found it difficult to understand why some terms were labeled archaic, some historical, some rare, when the same sorts

<sup>42</sup>The inconsistency and general unsatisfactoriness of editorial labeling in both OED1 and SOED was noted by Stein (1983, 468–75). See also Brewer (2003, 329–332), for additional discussion of Burchfield's practice, and Muggleston (2000) for an account of conflicting influences on labeling criteria in OED1.

<sup>43</sup>Here and there in the present volume I have found myself adding my own opinions about the acceptability of certain words or meanings in educated use. Users of the dictionary may or may not find these editorial comments diverting: they have been added (adapting a statement by John Ray in 1691) "as oil to preserve the mutelage from inspissation" (SOED 1972–1986, 3.v-vi).

of quotation evidence were adduced in each case. Passing my draft on to the editor, John Simpson, for comment, I was taken aback to learn that many of the inconsistencies had been already noted and corrected. He wrote back to me to explain that "soon after we started publishing online in March 2000 we recognized that there was a problem of consistency with our obsolete/rare labeling. This arose because the system we were using (developed from that in use on the Supplement) had too many unnecessary complications — and so the system was being applied slightly differently by different editors. As a result, in late 2000 we reviewed our policy and decided on a simpler approach. We are using this new policy now, and have applied the changes to much of the online text (as part of an ongoing procedure). I suspect you conducted your research on these labels before we had started to implement the new policy."

The result is that OED3 has done, is in the process of doing, and is even re-doing, an enormous amount of tidying up of labels. For example, historical words supported with one quotation alone are now routinely described as rare, as well as obsolete, whereas before they were labeled obsolete but not necessarily rare (as in the case of *magnetical*, supported with a single seventeenth-century quotation, *magniscandate*, which had a single nineteenth-century quotation, *magniscandant*, two quotations dated 1843, and many others). Some of the apparent lapses from consistency turn out to be explicable according to rational principle. Thus, the answer to my query why *misincline* is labeled obsolete, but *misimprison* labeled both obsolete and rare, when both are illustrated by a single 17<sup>th</sup> century quotation, was as follows: "It would have been easy to label both as obsolete and rare, but in fact *misincline* is supported by other early evidence (see *misinclined* and *misinclination*), whereas *misimprison* appears to be isolated. On the basis of this wider evidence it seems reasonable to withhold the rare label from *misincline*." This is a defensible if not transparent policy.

However, returning to the revised portions of the OED in the wake of Simpson's response, I have again found it reasonably easy to spot inconsistencies. For example, the last quotation for *nealing* *n* is dated 1839, and that for *neckclothed* dated 1864, but neither is labeled obsolete, whereas other words last quoted in the 1870s ARE labeled obsolete (as *necking* *n*<sup>1</sup>, *neckbreak* *adv*, and many other examples); *neckland* is labeled obsolete but not rare, despite having only two quotations (1598 and 1627), whereas other pre-1870 words with only two quotations are described as obsolete AND rare, etc., etc.). Individually, these examples and others are trivial; cumulatively, however, they may



be important, given that, as already described, labels are one of the discriminating factors by which the OED's wealth of material can be searched. But it is impossible to test the current revisers' consistency in labeling by any method which is itself consistent: sophisticated as the improved search features introduced into the OED Online site in January 2002 are, they do not enable searches which reveal how New Edition labels compare with those of the previous editions in relation to quotation dating and quotation frequency, and nor do they allow one to search for labels as a separate category — i.e., separately from all the other material that appears in the "definitions" text. It is therefore impossible, or so I have found, to do more than comb through entry by entry, serendipitously happening on this or that.

The search facilities do, however, open up some promising avenues of investigation into the comparative use in one edition or another of labels and labeling, and consequently into what this suggests about the various sorts of criteria applied by the lexicographers. Thus one can get a list of the occasions on which "Now rare" appears in the "definitions" text of OED2 entries, and count those instances occurring within (for example) the *M-momnisher* stretch of revised material — the answer is 144. The same query applied to OED3 yields 941 results, a significant increase. It may be that the new application of this label identifies words in regular use in the late-nineteenth century, but now on their way to obsolescence, and that searching for this label in OED3 is a potentially valuable tool for various sorts of lexical research. But to investigate OED3's use of the term in any more detail you would need to go through each of those 144 and 941 entries, check that "now rare" was indeed being used as a label, and compare OED2 with OED3, looking at the date and range of quotations, together with any other editorial comment, in each case.

Another promising label to explore (since it may tell us something valuable about any changes either in the revisers' choice of sources, or their judgements about particular usages) is "literary." The word *literary* occurs 111 times in the "definitions" text of OED3 (i.e., over the range of words *M-momnisher*), but only 50 times in the corresponding OED2 text — quite a big difference, and worth looking at further (have the revisers chosen to label as "literary" usages previously unlabeled? Or have they simply identified additional "literary" senses of words, or added more "literary" quotations?) "consciously literary" seems to be a new descriptive term in OED3, and could be very useful, given the dictionary's predilection to date for unusual, writerly diction of one sort another, although the phrase has been used just once: the

revisers say of *make v*<sup>1</sup> 'compose, write' (s.v. 4a), that "The principal modern use is of poems or verses, though even this is somewhat *arch*, or consciously literary."

Distinguishing between "archaic" (not "arch") and "consciously literary" as they do here seems helpful, but may be specious. We need to know how these terms are understood by the lexicographers (and others like them, for example, "[not consciously] literary," "poetic," both of which may shade into "obsolete," and/or "historical," and perhaps also deserve the addition of "(now) rare"). It is hard to see how the long-due overhaul of editorial labels can be carried out successfully if the lexicographers do not publish and explain their criteria for assigning labels, and supply a comprehensive list. At present, some labels appear in the list of abbreviations available both in print and online — for example, "colloq.," "derog.," "vulg.," "arch.," "hist.," "obs.," "poet." — but others, not being abbreviations, do not — for example, "affected," "coarse," "coarse slang," "emotional feminine," "humorously pedantic," "illiterate," "improper," "low," "low colloq.," "judicious," "now rare," "rare," "shoppy," "well known," and a number of other terms.

Various of these terms, unsurprisingly, have been dropped in the current revision, reflecting as they do attitudes or social judgments now out of date (though they still have value as indicating the connotations, of whatever sort, which words may have had in the past). Perhaps the most striking instance I have found of label change in OED3 is the revisers' virtually complete eschewal (so far) of the terms "erron." (i.e., "erroneous"), and "catachr." (i.e., "catachrestic"), as applied to usage, both of which occur many times in previous editions of the dictionary to indicate "incorrect" or contentious usage of one sort or another. The avoidance of such terms of condemnation points to a significant change in lexicographical position. All the OED lexicographers, from Trench onwards, have paid lip service to the ideal (variously stated) of descriptivism, but the OED3 editors are the first to prefer description to prescription in (almost) uniform practice as well as theory. They have largely turned their back on Burchfield's Fowlerian *ipse dixit*, as on OED1's more covert proscriptions of one form or another.<sup>44</sup>

<sup>44</sup>Trench (1860) famously envisaged the OED's job as the construction of an objective "inventory" of language; Burchfield many times declares (and often observes) the importance of descriptive lexicography, for instance, in the inclusion of sexual words and terms of racial abuse (see 1989, 109–15, for example). His *ipse dixit*, which run contrary to this policy, were variously dealt with



These few examples give some idea of the way in which editorial labels can tell us about the lexicographers as well as the words they describe. In addition — if the search engines can be persuaded to divulge this information — labels can reveal a significant picture of the types of sources the lexicographers are choosing to document, and hence the quality and function of the OED as a whole. Transparency, comprehensiveness, and consistency in both label explanation and label application are therefore vital; and this is one of the areas where OED3 might with profit consider extending and reviewing its policy of revision.

### Quotation Numbers and Distribution

For similar reasons, quotation distribution, as well as provenance, is another key area in OED study. We seek to understand the dictionary by scrutinizing editorial comment in relation to the style and content of quotations on the one hand, and their number and chronological spacing on the other. In particular, different numbers of quotations for different words often look striking in electronic searches of various kinds. But odd gaps in documentation are not unusual. They are often found in the countless entries where no twentieth-century documentation was supplied by Burchfield, and where there was consequently none in OED2, but where the Third Edition revisers, as observed above, have sought to update the record. Thus *misentry* has no quotation between Blackstone in 1768 and the *Economic Journal* in 1957, though it is instanced again in 2001, *majorship* has no quotation between 1875 and 2002 (and the 2002 quotation should probably be labeled "historical"), *maiden-like* (adj.) none between 1865 (George Meredith) and 1990 (*Stereo Review* — said of a soprano voice). In the absence of editorial comment on these documentary chasms, it is difficult to know what to infer from them.

by the OED2 lexicographers. Some (for example, the proscription of *opinion-nam*, or *permanence*) were retained in OED2 but tagged "R. W. B."; this label was nowhere explained but evidently designated Burchfield, and appeared to identify the judgment as idiosyncratic. Other of Burchfield's prescriptive comments (as on *agenda* and *layperson*) were dropped; an apparently smaller number (as on *disinterested* and *hopefully*) were reproduced without comment. See further Brewer (2005).

But how reasonable, or realistic, is it to infer anything at all? Any user of OED1 will be familiar with chronological unevenness in the distribution of quotations between and within entries, and will have put it down to quite understandable variations in the evidence available to Murray and his editors. On the one hand space was at a premium; on the other, the lexicographers have always had to limit the time they could afford to spend chasing up one particular instance of usage or another.<sup>45</sup> It was early apparent that quotations had to be kept down to about one a century per word or sense if the dictionary were ever to be completed, and the "General Explanations" printed in OED1 (and reproduced both in OED2 [1989, 1. xxix] and online) state that quotations are "arranged chronologically, so as to give about one for each century, though various considerations often render a larger number necessary." This means that the quotations "have also to illustrate . . . by negative evidence, [a word's] non-existence at the given date." Attaching significance of any sort to "negative evidence" is risky, though. There are remarkable variations in the number and chronological distribution of quotations supplied for words and senses. These appear to be much greater in the case of material supplied by Burchfield than in the First Edition (although it is impossible to check this impression in any systematic way). Thus, in his updating of the first edition's entry for the noun *alibi* (originally published in Murray's first fascicle of OED, in 1884), Burchfield supplied 2 extra quotations, one antedating and one a postdating, so that the new quotation range (found also, of course, in OED2), reads 1743, 1774, 1855, 1862, 1939 (this last from Eliot's *Old Possum's Practical Cats*). Useful, but a curious contrast to his treatment of the newly identified transitive verb *alibi* 'clear by an excuse, provide an alibi for', whose first appearance he dated 1909. Here, he printed not one, or two, but instead seven quotations, the first one of 1909, and then 1917 (two from same source), 1926 (two from same source), 1930, 1958. Whatever one should infer from this, the most obvious inference, that the verb *alibi* is either more frequently used or in some way more important than the

<sup>45</sup>Dictionary records — whether the OED archives at OUP, the Murray papers in the Bodleian library, or the various articles by lexicographers describing their labours — are full of references to publishers insisting on reducing the length of individual entries, and trying to restrain staff from the "drift to Bodley" to search for more quotations.

noun, must, surely, be erroneous.<sup>46</sup> Instead, this must be an example of the OED's habit of favoring eccentric diction at the expense of the core lexicon, something deplored by Murray from the early days of his editorship but still, evidently, hard to avoid.

Even more striking are the 12 twentieth-century quotations Burchfield supplied for *mantra*, nine of which come from the years 1962–1973, although according to Burchfield's normal rule the word did not need updating anyway since it was already furnished with nineteenth-century quotations.<sup>47</sup> Contrast that with the three quotations for *manufacturer* (1752, 1832, and 1901), three quotations for *Labour Government* (1926, 1945, 1971), two quotations (both eighteenth-century), for *labour-pains*, and two twentieth-century quotations for *rape* meaning sexual assault — both of which, incidentally, refer to male rape. Examples of this sort of variation, and the consequent puzzle as to how it should be interpreted or what it is intended to imply, can easily be multiplied. Burchfield's favorite authors tended to be rewarded in what appeared to be disproportionate ways. Thus he gave Woolf's coinage *scrolloping* six quotations (duly reproduced in OED2), describing it as a "Fanciful portmanteau formation . . . prob. combining SCROLL n., LOLLOP v., etc." She is the only quoted author for this word, as she is likewise for *vagulate* (three quotations).<sup>48</sup> The adjective *runcible*, which was not included in the original OED, was given 10 quotations by Burchfield, running from 1871 to 1979. Six of these are taken from the works of Edward Lear, who is identified as the first user of this nonsense word.

It seems churlish to object to such lavish provision, but the consequent unevenness in documentation — which the electrification of the dictionary lays open for the first time — provokes awkward

<sup>46</sup>Burchfield's choice of T. S. Eliot as his only source for twentieth-century use of the noun might well give the wrong impression that it is or was in some way an unusual, poetic, or fanciful term. Presumably it would have been easy to turn up dozens of examples from other twentieth-century sources—detective fiction for example, in which the prolific OED reader Marghanita Laski was an expert, and from which she could doubtless have readily furnished quotations (see Laski [1968]).

<sup>47</sup>Nor have we added later examples to words and senses whose illustration ends in the [first edition of the] Dictionary with nineteenth-century examples" (Burchfield 1972–1986, 1.xv).

<sup>48</sup>On Virginia Woolf and the OED, see Fowler (2002).

questions. It is hard not to draw the conclusion that, on some level at least, Burchfield believes that *scrollop* and *runcible*, whether in their own right as words or on account of their provenance, are more important than *rape* and the other words less generously attested. What defence could be mounted for illustrating the former at such great length, and the latter so minimally, given the lexicographers' claim that, as "an irreplaceable part of English culture," the OED "not only provides an important record of the evolution of our language, but also documents the continuing development of our society"?<sup>49</sup> The playful fascination of writers, whether English or not, with nonsense and other sorts of words could be argued to be an important index of the "culture" of the language they speak — as also Burchfield's almost equally generous documentation of scientific and technical terminology from a wide range of disciplines. But similar claims could be made for other aspects of "English" culture which Burchfield neglects (for some of these, see Strang [1974 and 1977]).

I have been able to find no discussion by Burchfield of quotation frequency and chronological distribution, as found either in SOED or in the first OED. It occasionally looks as if his general policy was to supply at least five examples of a new word,<sup>50</sup> but to be content

<sup>49</sup>Quoted from <http://www.oed.com/public/inside/history.htm#edrom>. Burchfield several times alludes to his treatment of literary usage, usually defensively, and claims that his literary quotations are "golden specks" by which "the balance of the volumes has not been disturbed" (see for example, Burchfield ([1989, 12]) and SOED ([1972–1986, 4.xi]). There are many occasions where he chooses literary quotations even when these shed only questionable light on the meaning of the word they are supposed to illustrate: thus Auden's "apotropically scowling, a tinker Shuffles past," is included as the last quotation, dated 1956, for *apotropaic* (rather than the 1966 example that still survives in the bundle of slips from which he made his choice, preserved in the OUP archives); Blinden's "tender amaranthine domes Of angel-evenings" is given as an example of the attributive use of *angel*; Auden's "How will you answer when from their quaking spring The immortal nymphs fly shrieking" is quoted (without comment) as the sole example of *quaking ppl.* a since Milton. To understand these quotations you need advance knowledge of the definition supplied by the dictionary: in other words the quotations depend on, rather than support, the definition.

<sup>50</sup>Some random examples: *perlocution* 'speech act': 6 quotations; *periphrase*: 7 quotations; *periphrastic*: 6 quotations; the verb *compere*: 5 quotations; *crappy*: 9 quotations (3 from an American thesaurus of slang); and very many others. With these, contrast *crampet* 'wallhook': 2 quotations; *crash* meaning the name of a tint in textile fabrics: 2 quotations; *creditable* 'capable



with far fewer — or indeed none — where there was nineteenth-century evidence in OED1 (as we have seen, there are countless examples of common words with no twentieth-century attestation whatsoever). However, there are too many instances of inconsistencies to be sure. And whatever Burchfield's policy, the result is that, where twentieth-century quotation evidence exists for a word or sense, it is often disproportionately large in comparison with the number of quotations for previous centuries — this despite the fact that, overall, quotation representation in OED2 for the twentieth century is lower than that for the nineteenth century.<sup>51</sup>

What then is the policy of the current OED revisers? Simpson, in the preface to the Third Edition, is a little more forthcoming than Burchfield. "Various factors contribute to the number of quotations that are used to illustrate the history of a particular word or meaning in the Dictionary," he writes in the section on "Documentation" in his online preface to the Third Edition. "In some cases (depending on the length of time a term has been recorded in English) an interval of fifty years between quotations might be appropriate. In others, a longer or shorter time span might be satisfactory. Other significant factors include the relative frequency of the term in a given period, the availability of quotation material, and the need to illustrate numerous spelling variants and grammatical structures."<sup>52</sup>

Such frankness is welcome, but this list of "significant factors" raises more questions than it answers. If "relative frequency of [a] term in a given period" affects the number of quotations with which the term is supplied, then it might seem reasonable to assume that more quotations for a word means that it was relatively common, and fewer quotations means that it was relatively rare. But such an assumption would be clearly false in the case of Burchfield's munificent illus-

of being assigned to: 1 quotation. Marghanita Laski, the single most productive contributor to Burchfield's supplement, stated that "five examples [were] wanted" of a new word, phrase, or meaning, if it was to be included in the dictionary, "though probably not more than three or fewer will be used" (1968, 38).

<sup>51</sup>According to the search tools on OED Online, there are 505,731 quotations from the period 1900–1999 in OED2, compared with 749,718 for 1800–1899 — and compared with 271,208 for 1700–1799, 380,610 for 1600–1699, and 246,135 for 1500–1599.

<sup>52</sup>See <[http://www.oed.com/public/guide/preface\\_3.htm#documentation](http://www.oed.com/public/guide/preface_3.htm#documentation)>.

tration of *mantra*, *scrolling*, *et al.*, as against the meagre representation of *rape* and of hundreds, probably thousands of other undeniably common words in OED1 and OED2.<sup>53</sup> Moreover, the "relative frequency" factor will on occasion be at odds with that of the "availability of quotation material," which Simpson puts next in his list. For while it may initially appear that finding illustrative quotations for common words will be easy, and finding illustrative quotations for rare words will be hard, such has not always been the case: Murray often complained that OED readers were far more likely to note down and submit unusual rather than usual words, with the result that "good quotations for common words were painfully deficient," and presumably Burchfield found the same (see J. A. H. Murray [1884, 516] and K. M. E. Murray [1977, 200–201]). (It is different for the current lexicographers, who can draw on the search facilities of databases which will not exhibit the same unconscious biases as individual readers). This may explain the relative scantiness of evidence in OED2 for the noun *alibi* as against the verb, and also (to choose randomly from many possible examples) *dialling* (vbl. n. s.v. 2b, 2 quotations), *dialogue* (s.v. 1a, five quotations between 1401 and 1865) and *meander* (used intransitively of a river, 3 quotations between 1612 and 1894) — all, one might assume, relatively common words, certainly more so than many others for which there is more quotation evidence.

Interestingly enough, the factor that Simpson doesn't mention is pressure of space. Yet this must have been one of the most crucial influences on quotation numbers and frequency for the first OED. The first editors were hounded by OUP to produce copy at a decent rate and to keep the quantity of material published — both the number of entries and their documentation — to as low a limit as was consonant with the OED's original aim (that is, of being the first dictionary "on historical principles" with quotations that illustrated the history of each word and sense treated). With the introduction of the electronic medium, the procrustean imperative, hitherto dominating dictionary

<sup>53</sup>There are other indications in SOED that Burchfield did not set out to correlate "relative frequency of a term" with number of quotations supplied for it. Thus he gives us four examples of the word *self-deliverance* 'suicide', all between 1975 and 1980, while telling us this euphemism "is not yet (1982) in wide currency." This is the same as the number documenting *serendipity*, a word which he states has had "wide currency in the 20<sup>th</sup> century", yet he offers only four post-1959 examples of *top people*, an "expression [which] gained wide currency from the advertising slogan used by *The Times* in 1957."

production, has been reduced, if not altogether removed (the editors have not altogether ruled out a further printed edition).<sup>54</sup> Column inches simply do not carry the same cost in cyber-space as on the printed page. The consequence, presumably, is that one might expect there to be more quotations in the third edition than in the preceding ones.

However, this seems not to be the case — or at any rate, not consistently so. I cannot claim that the following remarks are the result of a systematic examination of the evidence, because I cannot think how to manipulate the search tools so as to conduct such an examination. Incidentally, this illustrates one of the disadvantages of the electronic screen. One can swiftly flick through a hundred pages of print and gauge with the eye how many quotations there are, very roughly speaking, per entry, and to what period any bunching of quotation dates belongs; well illustrated words immediately stand out in comparison with sparsely illustrated ones, and idle browsing fruitfully indicates possibly rich lines of investigation. It is impossible to do any of this on screen, since one can only see a small amount of material at a time, and one cannot get a sense of the comparative length and character of individual or successive OED entries. Nevertheless, my impression is that, where the third edition revisers are dealing with material (largely) untouched by Burchfield, they provide a reasonably even spread of chronological attestation; in other words, they keep to the same rough proportions per century as the first edition did before them: for instance, see the entries for **magic**, **manipulate**, **mannerless**, **mannerly**, **mannish**, and many others. But where they take on an entry for which Burchfield has already supplied numerous contemporary

<sup>54</sup>Initially, Murray had agreed with the OUP to aim for a ratio of 6 printed columns of OED copy to 1 of The Merriam-Webster *American Dictionary of the English Language* (1864), but, as time went on, the lexicographers exceeded this limit by a greater and greater extent, so that the fascicles produced in the 1920s, towards the end of the alphabet, were as much as 16 times and more as long as the Merriam-Webster treatment of the equivalent material, eliciting fury and despair from Chapman and Sisam (as evidenced in numerous internal memos in the OED archives). We know that Burchfield was originally contracted (on Sisam's advice) to produce one volume only for the twentieth century supplement, and that this eventually swelled to 4 volumes; presumably, he was beset by the same pressure to produce, both quickly and briefly, as were his predecessors. Burchfield himself says tactfully, "it is no secret that the financial guardians of publishing houses still keep a stern eye on the waywardness and procrastination of their resident lexicographers" (1989, 192).

quotations, their treatment varies between near-complete preservation of his evidence on some occasions, and ruthless pruning on others (for example, removing 4 of his twentieth-century quotations for **Mahabharata**, 5 of his twentieth-century quotations for **madam** — in the sense 'brothel-keeper', 6 of his twentieth-century quotations for the adjectival use of **mogul**, and 10 of his quotations for **media**.)

But this leads to real imbalances in the new edition. A few examples: **Mercalli** (as in the scale applied to earthquakes) was a new word identified by Burchfield and supplied by him with 4 quotations dated between 1921 and 1923. OED3 recasts the entry to distinguish between an "attributive" and "absolute" sense, and furnishes a further 7 quotations, extending the date range to 1999 and bringing the total number of quotations to 11.<sup>55</sup> This is the same number of twentieth-century quotations as **mercenary** now has in the new edition, in its various 5 senses as adjective and noun, though it must, surely be a much commoner word.<sup>56</sup> and compares with 12 quotations for **manufacturer** distributed over its entire history, 1698–1984, with only four from the twentieth century (in its most usual sense, s.v. 2; OED3 has expanded OED1's quotations threefold). Burchfield's 12 quotations for **mantra** have been culled to 6, five discarded and one redistributed under a different category,<sup>57</sup> and many other quotations have also been dropped from other entries, although on principles that it is not always easy to divine. (Possible explanations immediately occurring — that the dropped quotations are less revealing than others, or that a more even chronological range of quotations is achieved — look implausible in a number of instances when one compares the quotations and finds them equally, or less, illustrative of the word, and when one sees that the new chronological range is no more evenly distributed than the first. Examples may be readily found in many of the revised sections of the verb **make**).

<sup>55</sup>OED3 identifies an additional absolute usage of the word (unmodified by "scale") but that still does not explain why so many examples of its usage over so comparatively short a time-scale were thought necessary.

<sup>56</sup>**Mercenary** was treated by OED1 but left untouched by Burchfield, and hence appears in OED2 with no quotations later than 1871.

<sup>57</sup>The revisers have identified an additional sense, 'A constantly or monotonously repeated phrase or sentence; a characteristic formula or refrain; a by-word, slogan, or catchphrase', and have illustrated it with a further four quotations (all twentieth-century).

## Conclusion: The Importance Of Sources

Is it merely pedantic and ungrateful, not to mention mean-spirited, to cavil at this and the other sorts of inconsistency, or apparent inconsistency, in so magnificently conceived and executed a revision? Why should these details be important, given the extraordinary range and depth of new and supplementary evidence and information the OED's editors have given us? I think they are important and I shall try to spell out why. Most of my comments in this article have been on OED's choice and distribution of quotations, and on editorial labelling of words in relation to the quotation evidence provided. Both these things relate essentially to the way in which OED has selected from the sources available to it, how it has interpreted those sources, and how it represents them in the dictionary. As already stated, understanding and assessing OED's relationship with its sources is fundamentally important to weighing up its nature, function, and quality. We need to turn the dictionary inside out, to look at its proportions of and criteria for selection from sources as the first step in making a judgement about it, because only then can we gauge the likely value of what it tells us about the words it explains and defines.

The importance of the lexicographer's choice of sources is difficult to overstate. If a dictionary is to be representative of the language as a whole, something claimed for OED, then it has to treat words so as to take some note of their relative place in the lexicon. This means regarding sources as your main taxonomic factor, not words. Historically, however, OED has engaged with language at the level of words — each one seen as independent, worth catching and recording, with an independent history that can and should be traced backward and forward in as much detail as possible. But electrification — not just of OED itself, but also of the vast tracts of language now covered by electronic corpora — has facilitated a different form of analysis, one that need not replace OED's diachronic lexicographical method but that can valuably enhance it, by providing a range of synchronic information on the typical context of a word and its frequency of use. In other words, the contextual information now far more readily available on the place of individual words within the (or a) general lexicon makes it possible for Oxford lexicographers to bring their pre-Saussurean dictionary into the contemporary linguistic world.<sup>58</sup>

<sup>58</sup>See Henri Béjoint (2000), chapter 6, for a recent overview of the relationship between lexicographical traditions and linguistics, and Sidney Landau (2000), chapter 6, for recent dictionary use of electronic language corpora.

As we have seen, electrification enables a user to manipulate OED quotation data as never before, and thus set about building up a picture of the dictionary's comparative use of different sources and different periods. But why should one have to do this at all? If it is self-evident that a dictionary will reflect its sources, and that those sources must be selected, consciously or unconsciously, from a potentially enormous welter of language in ways that crucially determine the character of the dictionary, then we should be looking to the lexicographers themselves to list and analyze their sources, an undertaking that is feasible now as it never was with OED1 or either of the twentieth-century supplements. It would then be possible for them and for us to build up a picture of the linguistic and cultural hinterland from which the dictionary has been compiled and whose language it thus presents. It would also be possible for us to understand and evaluate better the current lexicographers' assertion already quoted that, "far more than a convenient place to look up words and their origins, the *Oxford English Dictionary* is an irreplaceable part of English culture. It not only provides an important record of the evolution of our language, but also documents the continuing development of our society. It is certain to continue in this role as we enter the new century."

This grand statement implies an objective weighing up and impartial reflection of language, something which may be impossible given the dangerously labile concepts with which it treats ("English"? "culture"? "language"?). Certainly it is hard to square with the OED's inconsistent variation in quotation number, distribution, and provenance — in other words, in its treatment of sources. If the lexicographers believe that the OED represents culture, however defined, and wish us to accept this proposition, then they must lay bare for us the connections they make between the dictionary and culture, which means telling us which sources they have used, and in what ways.

But it remains a major undertaking to examine the OED's sources, notwithstanding the electronic aids we are now able to use. The preface to the Third Edition is only slightly less elliptic than its predecessors in the information it provides: in the past, we are told, the dictionary was "criticized for its apparent reliance on literary texts to illustrate the development of the vocabulary of English over the centuries. A closer examination of earlier editions shows that this view has been overstated, though it is not entirely without foundation. The revised text makes use of many non-literary texts which were not available to the original Victorian readers and their immediate



successors.<sup>59</sup> But no bibliography is provided of the third edition's current coverage of sources, and the bibliography of the second edition (which lists thousands of separate items) cannot be analysed electronically, and gives no indication of the relative importance of each source — i.e., how many quotations each has furnished for the dictionary. This would have been an overwhelming task, of course — though providing invaluable information for the lexicographers themselves as well as for users, since it would give a fascinating picture of the nature and scope of the dictionary so far.

The present writer is engaged in a study of the treatment of literary sources in OED1, SOED, and OED3. The early results are interesting, if inconclusive. For example, Virginia Woolf's work has been intensively reread for the new edition: she is now quoted 125 times over the revised *M-Nipissing* range, compared with 18 times for the equivalent text in OED2. The representation of some other literary authors has also been greatly increased — Charlotte Brontë (108 quotations in the revised alphabet range, compared to 54 over the equivalent portion in OED2), Elizabeth Bowen (89 compared with 27), Larkin (34:4), Plath (36:12), Joyce (32:4:179). The augmentation of T. S. Eliot and W. H. Auden is less striking though still notable (respectively 87 quotations to 61, and 72 to 52), that of Heaney probably significant (8 quotations in OED3, compared with none for the equivalent stretch of text in OED2 — although much of Heaney's oeuvre has appeared since the publication of Burchfield's supplement). Nadine Gordimer now has 19 quotations compared with three in OED2, Doris Lessing 19 compared with six; Emily Dickinson, however, has no increased attestation (she is quoted just once in the revised range by both OED2 and OED3, for *martyr*, in a citation originally inserted by Burchfield, who quoted her 59 times in all).

As one continues with more-or-less arbitrary searches such as these, one is tempted to put together a picture of the reading choices of the lexicographers, and to ascribe to them political programs of one sort or another: more equal gender balance? more eighteenth-century evidence? more post-colonial documentation? fewer eccentric or nonce usages from (some) poets? But if the lexicographers are acting with such intentions, then they must spell them out clearly to the

user, enabling us to understand the characteristics both of the old OED and of the new. Or if increased documentation of an author is the by-product of using material from a particular electronic database, then it is important that we should have that information, too.

Random searches like those above may mean almost nothing when conducted in a piecemeal way. Setting them in a framework, and understanding how they fit into the OED's treatment of lexicon overall, are undertakings that require substantial labour, conceptual on the one hand and highly detailed on the other. It is profoundly to be hoped that the Third Edition revisers are themselves engaged in such reflection on their work and that they will in due course disclose in full the vital information we need about the three lexicographical criteria identified above: how and why sources for the OED have been selected, how they have been read and excerpted from, and how the information from them has been used to produce this major new revision of "the internet's biggest, most prestige-laden reference book."<sup>60</sup> In doing so they will be able better to understand, and explain to us, the true nature of this splendid accumulation of knowledge and information.

### References

- American Dictionary of The English Language*. 1864. Ed. Noah Porter and others. Boston: G. & C. Merriam Company.
- Anden, W. H. 1963. *The Dyer's Hand*. London: Faber & Faber.
- Baker, Peter S. 1988. "A Supplement to OED: Sc-Z." *Notes and Queries* 233: 148–153.
- Béjoint, Henri. 2000. *Modern Lexicography*. Oxford: Oxford UP.
- Brewer, Charlotte. 1993. "The Second Edition of the *Oxford English Dictionary*." *Review of English Studies* ns 44: 313–342.
- . 2000. "OED Sources." In *Lexicography and the OED: Pioneers in the Unmodden Forest*, edited by Lynda Mugglestone, 40–58. Oxford: Oxford UP.
- . 2005. "Authority and Personality: Usage Labels in The *Oxford English Dictionary*." *Transactions of the Philological Society* 103 (forthcoming).
- Burchfield, R. W. 1958. "O. E. D. A New Supplement." *The Periodical* 32 (no. 261): 229–231.
- . 1974. "The Treatment of Controversial Vocabulary in The *Oxford English Dictionary*." *Transactions of the Philological Society* 1973: 1–28.

<sup>59</sup>See <[http://www.oed.com/public/guide/preface\\_3.htm#documentation](http://www.oed.com/public/guide/preface_3.htm#documentation)>.

<sup>60</sup>This quotation is from *The Guardian*, reproduced on the first page of OED Online <<http://www.oed.com/public/welcome/>>.

- \_\_\_\_\_. 1975. "The Art of the Lexicographer." *Journal of the Royal Society of Arts* 123: 349-365.
- \_\_\_\_\_. 1989. *Unlocking the English Language*. London: Faber & Faber.
- Collins Cobuild English Dictionary. 1987. Ed. John Sinclair and others. London: HarperCollins.
- Eliot, T. S., and Desmond Hawkins. 1940 (28 November). "The Writer as Artist." *The Listener* 26 (no. 620): 773-774.
- Fowler, Rowena. 1998. "Robert Browning in the *Oxford English Dictionary*: A New Approach." *Studies in Philology* 95: 333-350.
- \_\_\_\_\_. 2002. "Virginia Woolf: Lexicographer." *English Language Notes* 39: 54-70.
- Harris, Roy. 1982. "The History Men." *Times Literary Supplement* (3 September): 935-936.
- Knowles, Elizabeth. 2000. "Making the *OED*: Readers and Editors. A Critical Survey." In *Lexicography and the OED: Pioneers in the Untraded Forest*, edited by Lynda Mugglestone, 22-39. Oxford: Oxford UP.
- Landau, Sidney I. 2000. *Dictionaries: The Art and Craft of Lexicography*. 2nd edition. Cambridge and New York: Cambridge UP.
- Laski, Marghanita. 1968. "Reading for *OED*." *Times Literary Supplement* (11 January): 37-39.
- Levitt, John. 1989. Letter to the Editors. *Times Literary Supplement* (May 18-25): 545.
- Longman Dictionary of the English Language. 1984. Ed. H. Gay and others. Harlow and London: Longman.
- Morrison, Blake, and Andrew Motion, eds. 1982. *The Penguin Book of Contemporary British Poetry*. Harmondsworth, Eng.: Penguin.
- Mugglestone, Lynda. 2000. "Labels Reconsidered: Objectivity and the *OED*." *Dictionaries* 21: 22-37.
- Murray, James A. H. 1857. *Proposal for a Publication of a New English Dictionary by the Philological Society*. London: Philological Society.
- \_\_\_\_\_. 1880. "Ninth Annual Address of the President to the Philological Society." *Transactions of The Philological Society 1880-1881*: 117-174.
- \_\_\_\_\_. 1884. "Thirteenth Annual Address of The President to The Philological Society." *Transactions of The Philological Society 1882-1884*: 501-531.
- Murray, K. M. E. 1977. *Caught in the Web of Words: James Murray and the Oxford English Dictionary*. New Haven: Yale UP.
- Onions, C. T. 1928. "How the Dictionary is Made." *The Periodical* 143: 15-17.
- Oxford English Dictionary*. 1933. 13 vols. Ed. J. A. H. Murray, H. Bradley, W. A. Craigie, and C. T. Onions. Oxford: Clarendon Press.
- Oxford English Dictionary. Second Edition*. 1989. 20 vols. Ed. J. A. Simpson and E. S. C. Weiner. Oxford: Clarendon Press.
- Oxford English Dictionary. Third Edition*. In progress. Ed. J. A. Simpson and others. Oxford: Clarendon Press.
- Oxford English Dictionary. A Brief Account. Sir James Murray In Memoriam*. 1916. Oxford: Clarendon Press.
- [Philological Society]. 1859. *Proposal for a Publication of a New English Dictionary by The Philological Society*. London: Trübner & Company.
- Pijnenburg, W., and F. de Tollenaere, eds. 1980. *Proceedings of the Second International Round Table Conference on Historical Lexicography*. Dordrecht: Foris.

- Raine, Craig. 1979. *A Martian Sends a Postcard Home*. Oxford: Oxford UP.
- Schäfer, Jürgen. 1980. *Documentation in the OED: Shakespeare as Nashe as Test Cases*. Oxford: Clarendon Press.
- Stanley, E. G. 1990. "The *Oxford English Dictionary* and Supplement: The Integrated Edition of 1989." *Review of English Studies* ns 41: 76-80.
- Stein, Gabriele. 1983. Review of *A Supplement to the Oxford English Dictionary*, ed. By R. W. Burchfield. Vol. III: O-Scz. Oxford: Clarendon Press, 1982. *Anglia* 10: 468-475.
- Strang, Barbara. 1974. "A Supplement to O. E. D.: A-G." *Notes and Queries* 219: 2-13.
- \_\_\_\_\_. 1977. "A Supplement to O. E. D.: H-N." *Notes and Queries* 222: 388-399.
- Supplement to the Oxford English Dictionary*. 1972-1986. 4 volumes. Ed. R. W. Burchfield and others. Oxford: Clarendon Press.
- Taylor, Dennis. 1993. *Hardy's Literary Language and Victorian Philology*. Oxford: Clarendon Press.
- Trench, Richard Chenevix. 1860. *On Some Deficiencies in Our English Dictionaries*. 2nd edition. London: John W. Parker.
- Willinsky, John. 1994. *Empire of Words: The Reign of the OED*. Princeton, NJ: Princeton UP.