The century that followed the publication of Noah Webster’s *Compendious Dictionary* (1806) was characterized by an unprecedented period of innovation and achievement in American lexicography. It produced a stream of fine works of continually evolving scope, including the early Webster and Worcester dictionaries, the 1864 *Webster’s Unabridged* which served as James Murray’s “yardstick” during the composition of the *New English Dictionary*, the expanded *Webster’s International Dictionary* of 1890, and the recognizably modern *Funk & Wagnalls’ Standard Dictionary* of 1894. *The Century Dictionary*, based on Annandale’s expanded edition of the *Imperial Dictionary*, itself based on an early edition of Webster’s, was the most ambitious lexicographical enterprise of them all. Lavishly funded by the publisher of the Century and St. Nicholas magazines and edited by the famous Sanskrit scholar, William Dwight Whitney, himself an alumnus of the seminal 1864 edition of Webster’s, it was first published serially in six volumes from 1889 till 1891, and was continually revised and expanded by Whitney’s assistant, Benjamin E. Smith, until 1911, by which point it had reached twelve volumes with almost 10,000 pages defining over 500,000 terms.

One measure of the increasing scope of lexicographical work during this period is the size of the editorial staff needed to produce these dictionaries. Webster and Worcester, like Johnson before them, could still produce dictionaries that were largely their own work and hence bore something of their personal stamp. By the time of *The Century Dictionary* this was no longer possible. As a glance at the period photograph of the Century office will demonstrate, Whitney had a staff larger than that of many more recent dictionaries. In addition to general editors, Whitney and Smith engaged a veritable “dream team” of distinguished specialist editors, such as the botanist, Frank Hall Knowlton, the geologist, Frank Ward Lester, the
mineralogist, Edward Salisbury Dana of Yale, the philosopher, Charles Sanders Peirce, who contributed definitions for terms in such fields as astronomy, philosophy, mathematics, and mechanics. and the ornithologist, Elliott Coues, who contributed about 40,000 definitions of biological and zoological terms. In addition to his other duties, Coues commissioned as many as 1,000 new illustrations for the dictionary from the artist, Ernest Thompson Seton, at $5 a pop.

Clearly the Century Company had spared no expense in the production of the dictionary and its continued maintenance over a period of twenty years. It is something of a mystery, therefore, that they never revised the dictionary after 1911 and never reprinted it after 1914. It was a clear commercial success at that point, and generally regarded, with the possible exception of the still half-finished New English Dictionary, as the most complete and authoritative dictionary of English. It commanded a price far and above that of the other great dictionaries of the day. In January 1914 a set of the Century Dictionary volumes was listed for sale in a classified ad for $32.50, in comparison with $5 for Funk & Wagnalls’ Standard Dictionary and $6 for Webster’s New International Dictionary.

There seems to have been no one cause for the Century’s demise, but rather an amalgam of accidental circumstances. Whitney was already in fragile health when the dictionary was first published and he died not long afterward in 1894. Much of the credit for the gloriously expanded text really belongs to his young assistant editor, Benjamin E. Smith. As the new chief editor, Smith, who guided the editing of the Cyclopedia and the supplementary volumes, seemed to promise many more years of editorial continuity. But on February 24th, 1913, he died at the relatively young age of 56. The New York Times obituary published the next day noted that “his illness dates from the completion of the Century Dictionary, when he suffered a breakdown due to overwork.” A little less than a year later the dictionary’s great printer, Theodore Low De Vinne, was also dead at the riper age of 86. In his obituary, he is quoted to the effect that the “typesetting machines” already used by “newspapers and other publications...were of little practical advantage in the best forms of bookmaking, where the work must not only be beautiful and artistic, but must also combine features that could not be accomplished on machines.” His son, Charles D. De Vinne, was at the time secretary of the company, which seemed to have a secure niche in the market for fine printing. Yet by the

---

1They were all dead within a couple of decades of the final revision of the dictionary. Coues died in 1899, Ward in 1913, Knowlton in 1926, and Dana in 1935.


3He was no relation, as was sometimes maliciously rumored, of the publisher, Roswell Smith.

4New York Times, Feb 17, 1914 pg. 11.
end of 1922 the De Vinne Press went out of business. James W. Bothwell, who
had taken over as president after Theodore De Vinne’s death, blamed the labor
unions and the lack of demand for high quality printing:

    Labor unions are absolutely prohibitive of fine work. Under their
rule the worker becomes a mere machine. We are no longer able to
get the type of men suited to our needs... There is no longer the great
demand for fine printing. Production and quantity are the watchwords
today.5

At about the same time, parenthetically, my grandfather was involved in sell-
ing off the printing division of Harper & Brothers, which had been De Vinne’s
most formidable rival in American letterpress printing. Clearly change was in the
air in the 1920’s printing industry. There was a move toward cutting costs through
what would now be called “outsourcing”, with smaller specialty companies taking
over the tasks that would have been done in house before. This may have favored
smaller projects, but did not auger well for a gargantuan letterpress project like
the Century Dictionary. The Century Magazine itself had lost its panache by the
late 1920’s, when it tried to reinvent itself as a quarterly without much success.

In 1927 the Century Company, soon to be merged with another venerable
American publishing house, D. Appleton and Co6, made one more attempt at
lexicography with the three volume abridgment known as the New Century Dic-
tionary. This work, based on the surviving materials of the original dictionary,
was edited by H. G. Emory and K. G. Brewster. It reprinted many of the original
cuts though it did not include updated versions, for example, a more modern illus-
tration for “automobile”. It updated certain entries7 and added some new words
but left out many others, as well as the majority of the fine quotations that distin-
guished the full scale dictionary, preferring to save space with made up examples
when possible. In general it was an inferior work that did not match the qual-
ity of Webster’s 2nd New International, with which it had now more directly to
compete.

The news about the Century Dictionary from the 1920’s through the 1940’s
was mainly composed of obituaries as the more famous contributors, one by one,
died off. Fewer and fewer articles contained the once ubiquitous clause “according
to the Century Dictionary”. But while it gradually became less and less visible

6The announcement was made in March 1933 and the merger completed by that summer.
7For instance, it conceded defeat to James Murray about the etymology of “cockney”.

Raising the Titanic: Prospects for Reviving the Century Dictionary
to the public eye, it was never forgotten by scholars or professional lexicographers. Notable among these was Clarence Barnhart, who used the materials of the Century Dictionary in the fine American College Dictionary (1947) and later the New Century Cyclopaedia of Names (1954), which updated the Century Cyclopaedia volumes, and the World Book Dictionary (1963). In more recent times, Robert Burchfield and the later editors of the OED have made the Century Dictionary, full of scientific entries and rich in “real world” quotations, one of their main checking texts. Richard Bailey has estimated that more than 2,000 OED entries owe something at least to the Century Dictionary. Apart from many citations of the Century Dictionary itself, I find that it is not at all uncommon to find that a quotation at the very heart of an OED comes directly from the corresponding entry in the Century.

While this lexicographic salvaging enabled a sort of spiritual survival of the Century, such an “afterlife” was still not a life, and without active revision, the Century itself began to feel increasingly out of date. There have been a number of efforts in recent years to attempt a more thorough-going revival of the Century Dictionary. Notable among these was a proposal by Laurence Urdang for a traditional revision aimed at print publication. The financial backing for such an enterprise could not be found. During the original editing cycle, it was estimated that the Century Company would lavish “upwards of a quarter of a million dollars” on the project. The eventual cost was over a million dollars. It is not hard to imagine why the expense of such a project — translated into the dollars of one hundred and more years later — would prove daunting for potential backers. It makes one appreciate the millions of pounds that Oxford University Press has poured into the current revision of the OED. This is a daring investment, even for a product with a continuous publishing history. A traditional revival of the Century Dictionary, with its editorial and publishing gap, is a riskier proposition.

A completely different approach to reviving the Century came from a colleague of mine at Bellcore, Robert Amsler, who attempted to scan and OCR the Century as the basis of an electronic dictionary serving the community of computational linguists. Amsler envisioned merging several electronic dictionaries to create one electronic “mega-dictionary” that could be consulted by artificial intelligence programs attempting various natural language processing tasks. This project foundered largely because of the poor quality of OCR at the time (the late 1980’s), which was thoroughly incapable of dealing with the kind of complex typesetting found in a dictionary like the Century. (This is also one reason OUP

---

8The Dial, June 1886, p. 47.
chose the more expensive route of keyboarding the OED.)

A little more than twenty years ago Richard Bailey speculated about what the dictionaries of the “next century” might be like. He was speaking at the time when the OED had just been captured in electronic form and it was uncertain what role computers might play in lexicography. The focus was still on the use of electronic corpora as aids to traditional editing and publishing. Bailey sensed the important possibilities of computer networks, but noted that “access” was still a difficult problem and that desk dictionaries at least would still be widely used in the foreseeable future.\footnote{See Bailey, Richard W. “Dictionaries of the next century.” Lexicography: An emerging international profession. Ed. Robert Ilson. Manchester, UK: Manchester Univ. Press, 1986. 123-37.} The development of computer technology since then, especially the appearance of the World Wide Web as a vehicle for electronic publishing, has changed the situation considerably. We proved, with the OED Online, that with electronic publication a huge dictionary could be delivered over the Web in multiple forms and used as easily as any collegiate.

I came to the Century Dictionary cautiously: through my work at the OED and as I looked on, with some amazement and some amusement, at Amsler’s computer experiments. My later experience with the DjVu format convinced me that a page image project could effectively get around the OCR problems he’d experienced. Page image projects, of course, were not without their own problems. The high resolution images required to make a scanned page of small type readable were awkward to work with at best until the advent of advanced compression technologies such as DjVu. But even with clumsy page image projects, when users have recourse to an actual image of a page, it takes natural precedence over the scanned text, and makes the latter’s imperfections of secondary importance. The Century Dictionary Online, which went live in the spring of 2001, was designed to demonstrate this, and to show, as the OED Online did also, that web publication fundamentally changed the equation when it came to publishing and using very large dictionaries.

The electronic Century Dictionary Project has in fact gone through three distinct publishing phases: online publication with free Web access, a CD-ROM version published last year in conjunction with Princeton Imaging at a nominal fee to recoup expenses, and a new “print on demand” version from Gorgias Press that will reproduce the 1911 print edition and may be sold on a subscription basis. All of these are bound to the original set of TIFF scans that I had done in 2001, and the first two allow searching of the original, uncorrected OCR text derived from those scans. None of these involve any revision of the original text, though
the very existence of that OCR text, which, though still not perfect, is generally much better than the Amsler experiments of ten years before, makes such revision a distinct possibility.

Any attempt to update the *Century Dictionary* requires that several major preliminary tasks be undertaken and completed, regardless of how the a revision might eventually be implemented. First, the existing text must be captured in a format that will enable systematic editing. The two separate alphabetical ranges of the original edition and the supplement must be combined and proofread. New words and senses must be identified and a standard scheme of editing developed.

The editing text should exist in some kind of structured or semi-structured format. This implies the use of a software-independent markup language, such as SGML, XML, or HTML, so that at least the major structural elements could be identified and any functional typesetting information preserved. There are two ways such a text might be captured: keyboarding or OCR. With the *OED*, the original “Murray” text along with Burchfield’s four volume *Supplement* were first keyboarded with minimal structural indicators, then converted to SGML, and then combined with a small set of new entries to form the text of the second edition. The text of the *Scottish National Dictionary*, on the other hand, was scanned and OCR’ed, saved first as Microsoft Word and then HTML to capture vital structural information, and then converted automatically to XML and proofread.[10] Neither data-capture method is error-free, and thus both require proofreading at some point. Keyboarding is relatively expensive, especially when the text has complex typesetting, but if done right, that is, with double or triple keying and SGML or XML markup, it has the advantage of being able to capture the major typeset-driven structures from the outset. Advances in OCR engines such as ABBY Finereader have made scanning and OCR a much more attractive option than it was even in the relatively recent past. A program like Finereader can recognize and capture most typesetting information and OCR texts can now be directly saved in a variety of formats, including HTML. But even the best OCR is still not perfect.

Recent tests conducted using “untrained” Finereader on the *Century Dictionary* were interesting in this respect. A text like the *Century Dictionary* is problematic for both human keyboarders and OCR programs. It contains text in many different languages that use different alphabets and different special characters or

[10]The text of the *Dictionary of the Older Scottish Tongue*, which has been published with the *SND* in the new *Dictionary of the Scots Language* online, had been partially keyboarded and converted to OUP-style SGML. The rest of the dictionary was captured from paper using the method of the *SND* and combined with this text.
symbols. There is also a special system of phonetic respelling, which includes, among other peculiarities, vowels with markings above and below the character. These accents were either lost or assigned to the wrong line of text. Without “training”, the Century’s daggers marking obsolete terms were routinely seen as small “t’s”. Finereader boasts of being able to read text in over one hundred thirty languages, but this feature cannot be used to full advantage in a dictionary text. I found that using English and Greek yielded good results for capturing text in the two main alphabets, but it lost most accented letters in English. Adding French and German to the mix yielded mixed but essentially negative results. The French and German accents were now recorded, but other accents, such as macrons, were now mis-recorded as umlauts, and diereses below a vowel were sometimes assigned as umlauts in a word on the next line. Even worse, the Spenserian spelling “hadde” was eagerly and ludicrously converted to the German “halbte”. And for certain dictionary abbreviations, Finereader, with its “dictionary lookup”, actually did worse than the older, character-based OCR of Expervision RTK. Another problem has to do with recognizing columns of text. Generally, Finereader did a very good job with this, but occasionally it was thrown off by the illustrations, which it tried to recognize separately and encode as images, and thus mangled the order of the entries on certain pages. This is not really a problem when the OCR text is subservient to the page image itself and is used mainly for searching, but it is a serious issue in creating an editing text.

This said, Finereader was remarkably accurate with the great majority of the text. Many of the errors that did occur happened because the OCR engine used for the test was untrained. The Finereader engine, however, can be trained with special patterns to recognize elements like the daggers and phonetic characters more accurately. Even in the test mode it was able to produce HTML files that looked, in one version and with certain browsers, almost eerily like the pages of the original dictionary: in three columns with embedded illustrations! Seductive as it was, this output format, loaded down with complicated display coding, was not the most satisfactory as a base editing text, but other, simpler versions promised something more useful. More important, the OCR text contained enough typesetting information that the main structures of the dictionary — the entries, headwords, and parts of speech — could be identified and coded automatically using simple postfilters.

A base text like this, whether captured through keyboarding or OCR is essential for the completion of two of the other main preliminary tasks: merging

---

11The choice may depend simply on how much money is available.
the two existing alphabetical ranges and identifying an initial target list of “new”
words to be added to the dictionary, both of which can be accomplished at least
partially through automated text processing. It should be noted that, up to this
point, there is almost no need for human editorial intervention. The resulting text,
subject to caveats about its imperfections and assuming recourse to the author-
ity of the page images of the original, could even serve as part of a public user
interface.

We now come to the million dollar question, however: how can an active
editorial revision actually proceed? There are several publishing models that could
be used in going forward with the Century Dictionary. One would be to take the
base text as the starting point for a traditional revision process along the lines of
the OED. The result might be a modern American dictionary second to none in its
richness of material and of considerable ongoing value to its publisher. The costs,
though less than a print driven revision, would still be considerable and would
require sustained financial backing. As I mentioned before, lexicography on this
scale is no longer the one man affair it was for Noah Webster. Even considering
the economies made possible by electronic production, storage, and distribution,
recruiting and training a staff of editors capable of undertaking a revision of the
Century Dictionary would require an “angel” with very deep pockets indeed. I
believe Bill Gates has already done a dictionary, and I don’t yet hear anyone from
Google knocking at the door.

An alternative approach might be go with some form of the “open source”
Wikipedia model, and make the text available for revision by Internet volunteers.
The result here, though harder to predict, could be a free and continually vital
Internet community dictionary. At first blush, the idea — not only of a new pub-
lishing model, but a new creation model — sounds heretical if not downright sub-
versive. I’ll confess that when I heard about Wikipedia in 2001 my first thought,
coming from a traditional publishing background, was that this was a horse that
would not run. Since that time I’ve been more and more impressed by what the
Wikipedia volunteers have accomplished. The looseness of its vetting mechanism
still makes me a bit queasy, but the self-correcting power of the Internet public has
produced work that is often of a very high quality — at least in the fields I know
something about — and I find myself sometimes choosing Wikipedia over my
long-held subscription to the Encyclopædia Britannica. For many of the young
students of my acquaintance, the choice is much easier to make.

But even if it is possible to create an encyclopedia with volunteers, surely
a dictionary requires “professional” lexicographers, or at least closely directed
lexicographers. A dictionary is different from an encyclopedia. It is more tightly
structured, and definition writing is more stylized and compressed than the prose suitable for encyclopedia articles. Who but a professional lexicographer, after all, can distinguish parts of speech, separate homonyms, or split senses? Well, perhaps an enthusiastic army of ordinary, intelligent people criticizing and correcting each other’s work as they do with the Wikipedia could do these things. Most of the lexicographers I know are, after all, ordinary, intelligent people who happen to have fallen into the work and who get better with practice. Some of them might even like to volunteer for such a project. The Internet has found ways, again and again, to expose professional exclusivity as just another form of hubris.

Perhaps the more telling argument against an open source Century Dictionary is not that it couldn’t be done, but that it shouldn’t be done. The Wikipedia is by definition a continuous work in progress, and thus always in a state of provisionally useful imperfection. Even a traditional revision would involve making major changes to what is, after all, a public icon of American scholarship, and such icons, as the experience of Webster’s Third New International and perhaps the OED3 will tell us, are not gladly tampered with, even when it is done with the blessing of their publishers. The reading audience comes to assume, rightly or not, a proprietary relation to such works, and any changes, however well intentioned or necessary, are not accepted lightly. For some people, it might seem best to leave the work as it was and is, a monument to past glory increasingly crusted over by the passage of time and of decreasing relevance. Whether the Century Dictionary were to be revised traditionally or untraditionally, the resulting work is bound to be different from the dictionary generations of readers have loved. Yet only if it is revised can it carry forward its manifest excellence into the reaches of the present century and perhaps beyond.