

John Simpson

Mediums of access to the *Oxford English Dictionary*

The *Oxford English Dictionary* (*OED*) was not the first of the great national historical dictionaries edited in the nineteenth century, but it has been one of the most successful. This paper examines how the contents of the dictionary can be manipulated computationally to offer the user new modes of access for investigating the language. These methods are of course relevant not only to dictionaries of the English language, and similar techniques are being followed by lexicographical scholars and computational linguists throughout the world.¹

1. The basics

A dictionary needs a tight editorial policy (covering the selection of entries, its defining and etymological methodology, etc.) and a formally established structure in which the lexical information is held.

The following illustration shows a typical single-sense *OED* entry for a borrowed word, in this case the word *puttony* (a kind of basket used in the wine trade) from Hungarian. A short examination of the entry reveals the marriage of content and structure. We need to understand the structure of the content before we can appreciate what is feasible computationally.

The **headword form** *puttony* shows us the dictionary spelling of the word, based on an analysis of the evidence of the word's usage available to the editors.

This is followed immediately by the **pronunciation**, offering a standardized British and American English variety. This isn't 'absolute' information, but is a guide to how the word is or might typically be pronounced in both regions. After this, there is **inflectional** information on ways in which (in this case) the plural form may be written. For a term which is still recognized as a foreign word, it is not surprising that both the Hungarian plural and a standardized English plural form is provided.

The **etymology** is more extensive, and is one of the features of any entry that presents additional hooks for computational analysis. In this case we see that the English word comes from modern Hungarian (we use the same spelling form in English), and that the Hungarian word derives from the regional German *Butten* recorded in Austria and Bavaria. Several additional notes supplement the detailed information of the word's form and pronunciation. The etymology is significant in reminding us of the historical links between Hungarian and German. A word would not pass from the southern German dialects into Hungarian without some reason, and of course geographical propinquity and common industrial and trade interests provide possible reasons.

¹ *OED Online* images are copyright to Oxford University Press and are reprinted by permission of the Secretary to the Delegates of Oxford University Press.

The **definition** is generally regarded as the heart of a dictionary entry, though in a national dictionary such as the *OED* it has to fight for this billing alongside several other major information sections. Here, we see that the definition immediately remarks that the sense is known principally in Hungary (so that readers are not misled into thinking it may have more widespread currency). The basket is defined concisely and appropriately, and its function is then briefly described. A secondary definition is added, arising from the documentary evidence of the quotation paragraph. This definition is not for a form of basket but a measure (a basketful) of dried grapes used in the manufacture of the wine Tokay.

puttony, *n.*

Text size: A A

View as: [Outline](#) | [Full entry](#)

Quotations: [Show all](#) | [Hide all](#)

Pronunciation: Brit. /'pu:tɒn(jə)/, U.S. /'putən(jə)/

Inflections: Plural *puttonyos, puttonos, puttonys, puttonyok*.

Etymology: < Hungarian *puttony* basket of wood or wickerwork, open at the top and becoming narrower and elongated towards the bottom, to be carried on the shoulder, now chiefly denoting a basket for gathering grapes < German regional (Austria, Bavaria) *Butten*, regional variant of *Bütte* basket (in southern Germany, Austria, and Switzerland also in unmutated form *Butte*): see *BIT* 71.³

The form *puttonyos* reflects the Hungarian adjective *puttonyos* 'holding as much as goes into one puttony' (in e.g. *öt puttonyos tokaji aszú* five-basket Tokay Aszú; < *puttony* + *-os*, suffix forming adjectives), rather than the Hungarian plural form (which is *puttonyok*, whence the occasional English plural *puttonyok*), but was reaprehended in English as a plural form, apparently due to its final *-s*.

O.E.D. Suppl. (1982) gives the non-naturalized pronunciation (pu-tɒn') /'putɒn/, which reflects the Hungarian pronunciation (in Hungarian, *ny* is the conventional spelling of the palatalized nasal /ɲ/).

(Show Less)

In Hungary: a basket usually made of wooden staves or wickerwork and designed to be carried on the back, used to transfer grapes from the vineyard to the wine press.

Hence: a measure of dried overripe grapes equivalent to the contents of such a container, added to the fermentation cask in making Tokay, the number of measures added determining the richness of the wine.

Thesaurus »
Categories »

1940 H. J. GROSSMAN *Guide to Wines, Spirits & Beers* viii. 99 The grapes are gathered in wooden vessels known as *puttonyos*, holding about 25 quarts.

1959 W. JAMES *Word-bk. Wine* 190 Pickers carry a small container called a puttony, into which they put selected overripe berries.; the number of full puttonyos mixed with the ripe but not overripe grapes in the fermentation cask determines the richness of the wine, the range being from one to five (and occasionally six) puttonyos.

1967 A. LICHINE *Encycl. Wines & Spirits* 426/2 The collar label on every bottle of Tokaji Aszu will state: 3 Puttonos, 4 Puttonos, etc.—the export agency simplifies the spelling by removing the 'y'.

1972 *Guardian* 26 Jan. 9/5 In theory there can be a Tokay of 6 puttonyos; but in practice only the five turns up.

1989 *Daily Tel.* 16 Sept. (Colour Suppl.) 66/3 Tokay Aszu—The more puttonys, the longer it can live.

2005 *Wine Internat.* Jan. 54/1 For this tasting we concentrated on sweet Aszú wines ranging from 4 to 6 puttonyos as well as the modern late-harvest styles.

(Hide quotations)

Fig. 1: The *OED Online* entry for *puttony n.*

“puttony, n.” *OED Online*. December 2012. Oxford University Press. 2 January 2013
www.oed.com/view/Entry/155284?rskey=SKKWxt&result=22&isAdvanced=true

Finally, the entry structure is completed by the paragraph of **illustrative quotations**. These document the occurrence and use of the word (or each sense of a word) in English, from the earliest appearance up to the date of editing (or of the term's obsolescence).

In the case of *puttony* we can see that it is at present first recorded in English in 1940, in the context of wine-making and wine-drinking, and it continues in this use into the 21st century, moving out of the highly specialized register of wine guides to the slightly less ethereal wine pages of the newspapers.

In multi-sense entries definition and quotation sections are repeated to accommodate each sense.

From this short explanation of a rather minor entry in the register of the English language it is possible to suggest the wealth of information classified in many of the much larger entries. With competent computational retrieval and display tools it should be possible to generate significant information about the language.

But this is not the only aim of the online dictionary. Historical dictionaries tell us not simply about words, but about the people who use them and the times and places in which they are used. If we question the dictionary properly, then we may be surprised at the amount of social and cultural data we can extract and analyse. Having said that, we should also bear in mind that the current software does not necessarily enable us to discover everything we would like to know from the data, and that future technological developments are likely to improve the quality of the information we can uncover from the dictionary's entries.

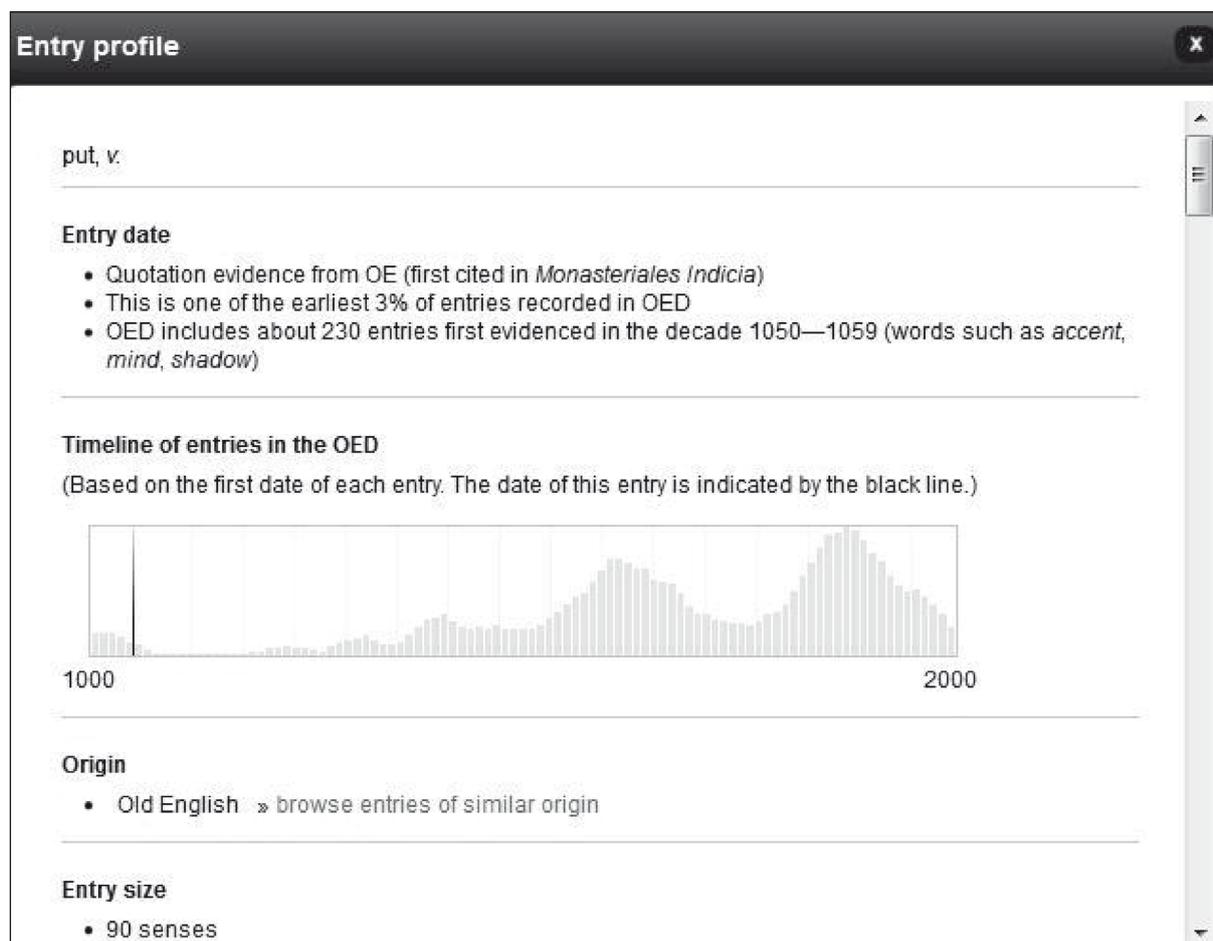


Fig. 2: Entry “profile” for *put* (OED Online)

1.1 An entry “profile”

Dictionaries can seem dry and unwieldy. One of the advantages of having access to a big dictionary online is that you can often choose “views” of the data with which you feel comfortable. Near to *puttony* we have the enormous entry for the verb *put* – one of the biggest entries in the whole of the *OED*. For every entry in the online *OED* there exists an entry “profile” generated from the data within the entry. If any feature of the entry changes when the database is regularly updated, then these changes will also be applied to the “profile”.

The entry “profile” for *put* (see Fig. 2) is quite extensive. It starts with some general information about the word: it dates from the Old English period, and is (according to the *OED*) one of the oldest 3% of words recorded in English. The *OED* also gives you access to other words first recorded in the same decade (1050-1059), such as *accent*, *mind*, and *shadow*. The word's position as one of the oldest words in the language is plotted on a timeline (the dark needle at the left).

Not shown here, the “profile” continues with a further digest of information: a visual representation of the emergence of new senses of *put* over time; lists of words referred to and referring to *put*; etc. By using computational means, we can select and display information which the reader requires before launching head first into a substantial entry.

1.2 Filtering results

One of today's computational idioms is the “filter”. This process of interrogation involves iteratively creating subsets of subsets until you arrive at the level you require.

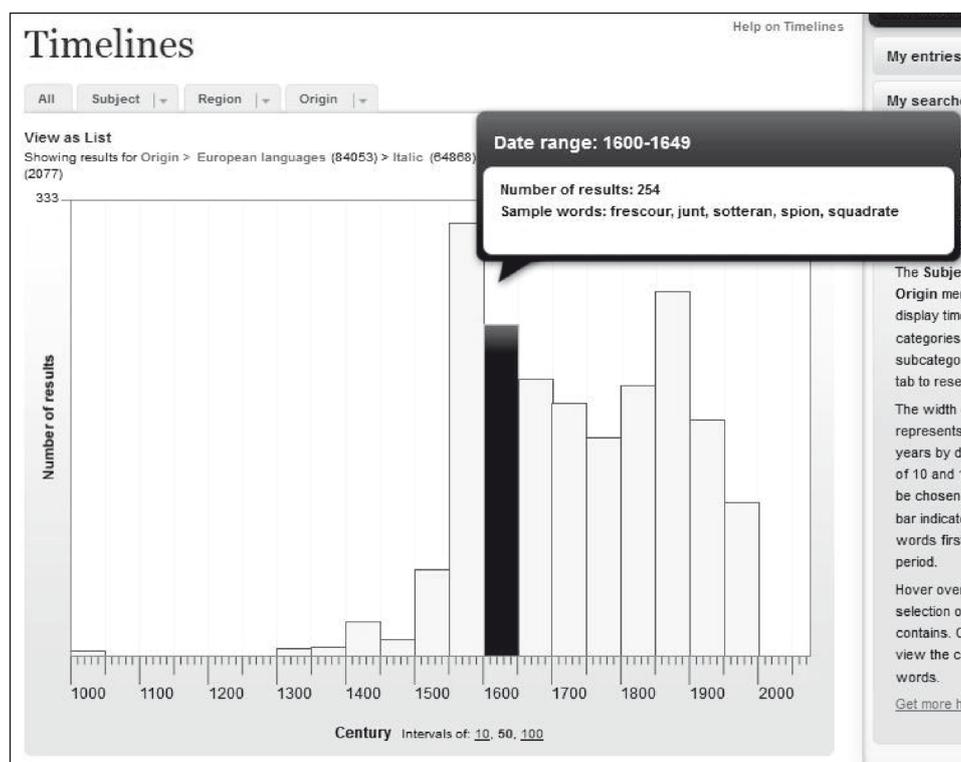


Fig. 3: Graph showing *OED* entries from Italian, arranged on a timeline based on their earliest reference in English

This means of access has been applied to numerous areas of the *OED Online*. We can take languages as an example. It is now possible to ask the database to report back on each of the entries which derive immediately from, say, Italian. Formerly this search would have included any entry which mentions the word “Italian” anywhere within the etymology text of an entry. But more recently the tagging of etymologies has been strengthened so that the software is able to disregard chance references to “Italian” and locate only those entries deriving directly from the language.

The results are revealing. Disregarding outliers (which are often problematic), the graph shows that although a number of Italian words entered English in the fifteenth and early sixteenth century, the first substantial influx occurred in the late sixteenth century. We can plot this against our real-world knowledge of when the influence of the European Renaissance started to have a substantial effect in Britain, and when we might expect the vocabulary of the Renaissance to begin to emerge in English texts.

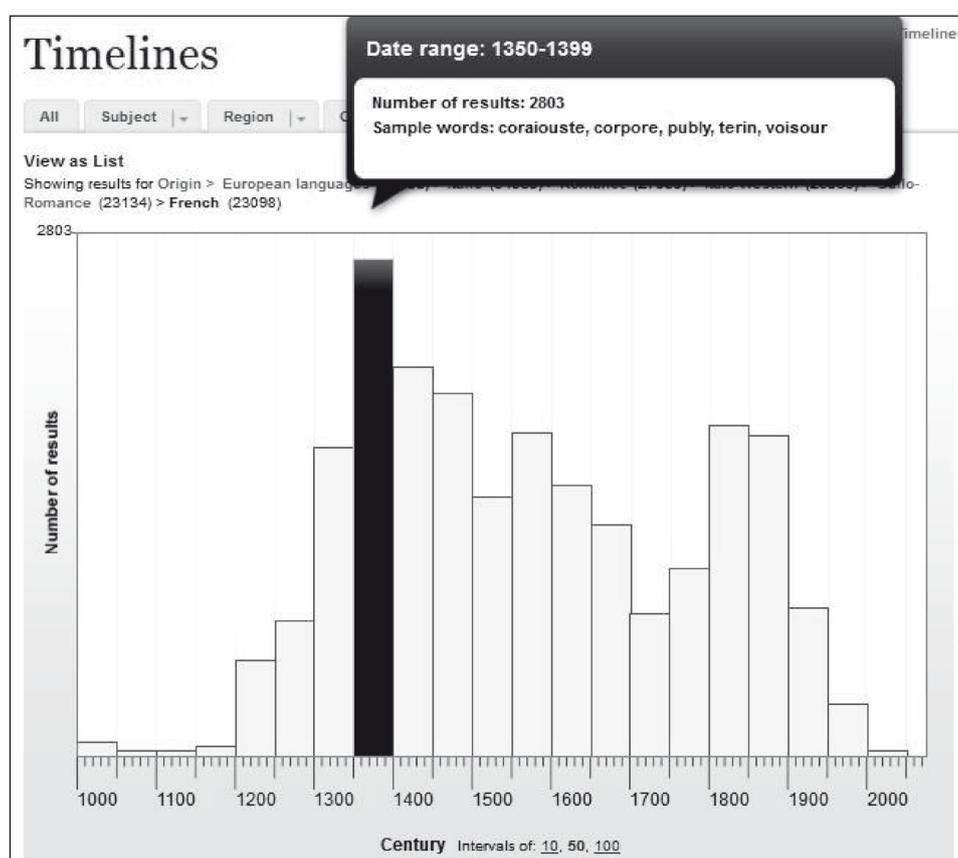


Fig. 4: Graph showing *OED* entries from French, arranged on a timeline based on their earliest reference in English

Note also that if the cursor is allowed to float over any of the columns of the graph, then a small box appears listing some of the Italian words which entered English in this time period (the span of which can be altered). A full click on any column will take the reader to a traditional-style word-list of all of the words from Italian which are recorded by the dictionary as entering English at this time. The possibilities for this sort of display in schools, for example, are easy to see. Students could be asked to compare the sorts of words which entered English from Italian (or any other language) in, say, the late six-

teenth century and the late nineteenth century. What do the different types of words illustrated in these two time periods tell us about the relationship between Britain (or the English-speaking world) and Italy then? What types of new vocabulary are characteristic of each period? And so on.

A wider picture can then be obtained by comparing the picture for Italian words with that of French words which entered English, as documented by the *OED*. The picture is, as you might expect, slightly different from the Italian picture. The shape of the graph is different, with an earlier peak for French in the late fourteenth century, documenting a different type of relationship between the two languages and the two countries. This is also hinted at by the bare numbers: although there are 2,077 Italian words recorded as the basis of English words, the number of French words is far higher, at 23,098. And the types of words introduced from Italian differ in kind from their French counterparts. Altogether, an analysis and comparison of each graph allows us insights into the relationship between the languages over time.

We might also note that a stronger picture would emerge if we had comparable information on Italian loans to and from French (and vice versa), to complete the triangle of languages. We can hope that in future more information may become available about the relationship between other languages.

The issue of filtering does not stop here. One can, for example, take the English words from French and request a timeline graph restricted simply to scientific (or political, or colloquial) vocabulary. Here is the graph for scientific terms:

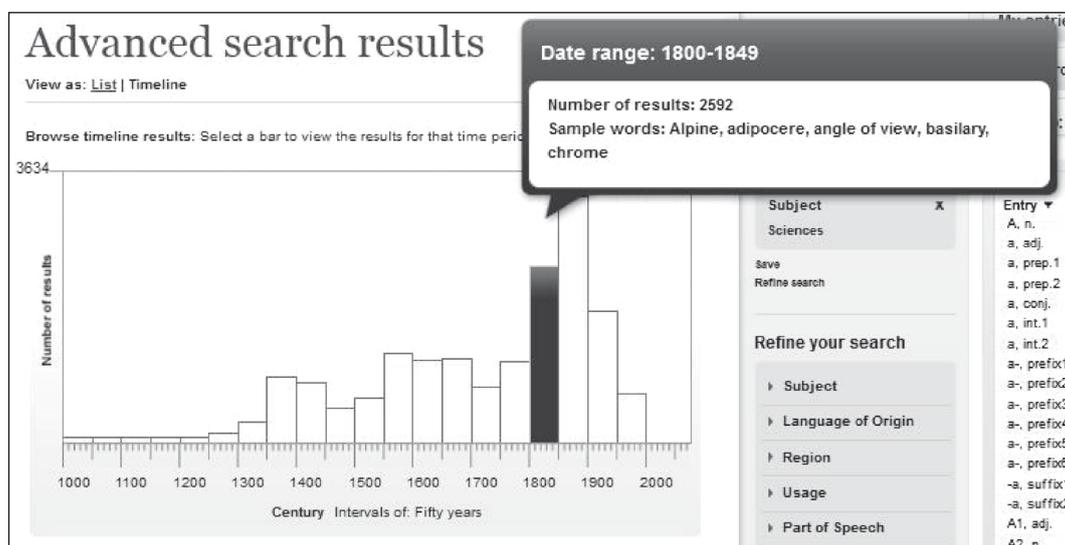


Fig. 5: Graph showing *OED* entries from French which are also categorized as scientific, arranged on a timeline based on their earliest reference in English

Here the peak has shifted to the late nineteenth century, but with a solid series of results from the late fourteenth showing a continuing involvement by the English in scientific words deriving from French.

The concept of filtering is not, as we have seen, restricted to language words. The filter can operate on any of the defined subject/discipline categories covered by the dictionary,

on geographical spread, on register (slang, colloquial usage, etc.), and on a number of other criteria. Overall, it forms a very powerful tool for accessing the historical data locked in the *OED*.

1.3 Taking trips outside the *OED*

Looking inside the *OED* is rewarding work. Nowadays the focus of interest is not only the individual word but, through advances in computational lexicography, the language as a whole. We can at last find out information about linguistic trends, if we ask the right questions.

But there is also a world outside the dictionary. The more external resources that the dictionary can link to, the stronger it becomes. In 2009 Oxford University Press published a *Historical Thesaurus of the OED*, which was the result of a long research programme at the University of Glasgow. The dominant idea behind the thesaurus was to produce a reference tool which allowed the user to search by theme rather than by word. It might enable one, for example, to search for all of the legal terminology covered by the *OED*.

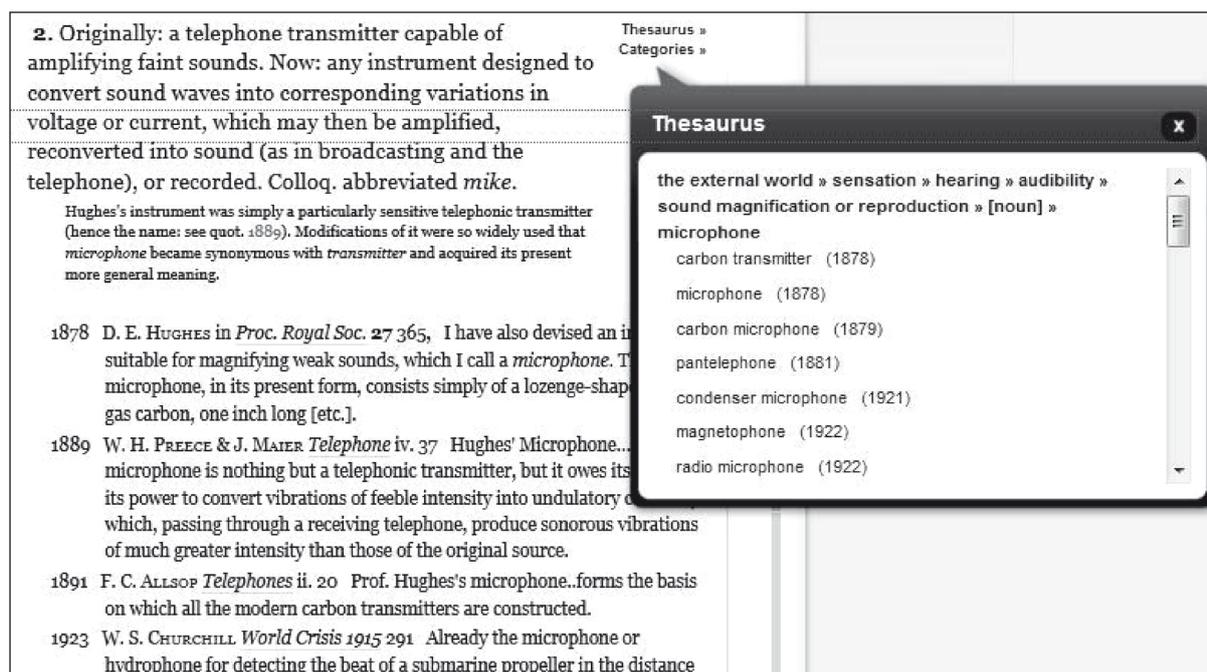


Fig. 6: Links from the *OED*'s entry for *microphone* to a chronological list of synonyms or near-synonyms classified by the *Historical Thesaurus of the OED*

The project sought to achieve this not by the simplistic route of searching for all entries and senses explicitly labelled “Law” (which would have missed many legal terms which have not been explicitly labelled for one reason or another) but by adopting a more systematic approach. The procedure was to write out almost all of the meanings of words in the *OED* on to index cards, along with their date of first use. These cards were then ordered thematically. The result is that the book allows us to find all words in the *OED* relating, for example, to motor transport.

One of the searches described above already uses this thesaurus, which has now been “plugged into” the *OED* computationally. We found all words of French origin in the dictionary, and then filtered that down to all scientific words of French origin. In order to isolate the ‘scientific’ words, the software used the thesaurus linkages.

These links can be seen at many entries throughout the dictionary. The illustration shows the *OED*'s entry for *microphone* with links displayed as targets elsewhere in the dictionary. The *Historical Thesaurus* is able to show that at much the same time as the word *microphone* was coined (1878), English-speakers were using several other terms for the same concept (in this case *carbon transmitter*, *carbon microphone*, and *pantelephone*). With the power of hindsight we can say how pleased we are that the right word won out!

The online site also allows readers access to the framework of the thesaurus, and then – in the process of following up a lead or suggestion – back to the entry in the dictionary itself.

<ul style="list-style-type: none"> └ punishment <ul style="list-style-type: none"> └ [noun] └ [adjective] └ [adverb] └ [verb (intransitive)] └ [verb (transitive)] └ [phrase] └ retributive punishment └ corporal punishment └ capital punishment └ torture └ imprisonment <ul style="list-style-type: none"> └ [noun] └ imprisoning [adjective] └ imprisoning [adverb] └ be imprisoned [verb (intransitive)] └ imprison [verb (transitive)] └ imprison [phrase] └ prison <ul style="list-style-type: none"> └ [noun] └ [adjective] └ [adverb] └ punish in prison [verb (transitive)] 	<p>lying-house 1423-4 (see quot. 1593).</p> <hr/> <p>javel 1483 = JAIL. <i>n.</i></p> <hr/> <p>toll-booth 1488 A town prison, a gaol.</p> <hr/> <p>kidcot(e) c1515 The name formerly given in various towns (as York, Lancaster, etc.) to the lock-up or prison.</p> <hr/> <p>clink c1530 The name of a noted prison in Southwark; later used elsewhere (esp. in Devon and Cornwall) for a small and dismal prison or prison-cell, a lock-up. Now used generally for: prison, cells.</p> <hr/> <p>warding-place 1573 (a) a place where guard is kept; (b) <i>Sc.</i> a place of confinement for prisoners.</p>
---	--

Fig. 7: An extract from the list of words associated with the thesaurus head *prison* on the *OED Online*

This snapshot of the framework shows (on the left) how “prison” fits into this section of the thesaurus structure: under “imprisonment” and in turn within “punishment”. The list on the right shows some of the senses identified by the thesaurus software as more or less synonymous with “prison”. This structure may be found throughout the entire dictionary, and therefore represents a significantly enhanced network of relationships available to the reader, in comparison to the cross-reference links which were the only form of linkage available previously.

It is now also possible to link outwards from an entry in the *OED* to equivalent entries in other dictionaries. This feature requires the user to have a subscription to the target text if one is necessary. So if the reader would like to find additional detailed information on a word current in the Middle English period, he or she would visit the relevant link in the central panel of the *OED*'s entry, and click through to the equivalent entry in the *Middle English Dictionary*. The same linking functionality is available with the *Dictionary of Old English* (in progress in Toronto), from where it is also possible to link back into the *OED*. An extract from the *OED*'s entry for *god* shows the links currently available outside the dictionary to other reference texts:

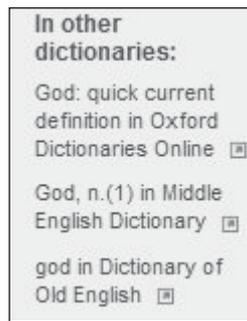


Fig. 8: Link from OED entry to other reference sources

It would be a relatively simple matter to engineer links to other texts: to other major on-line dictionaries, to the texts of sources cited as quotations in the dictionary, and to other data which might be useful to the reader when investigating a particular word or trend in language.

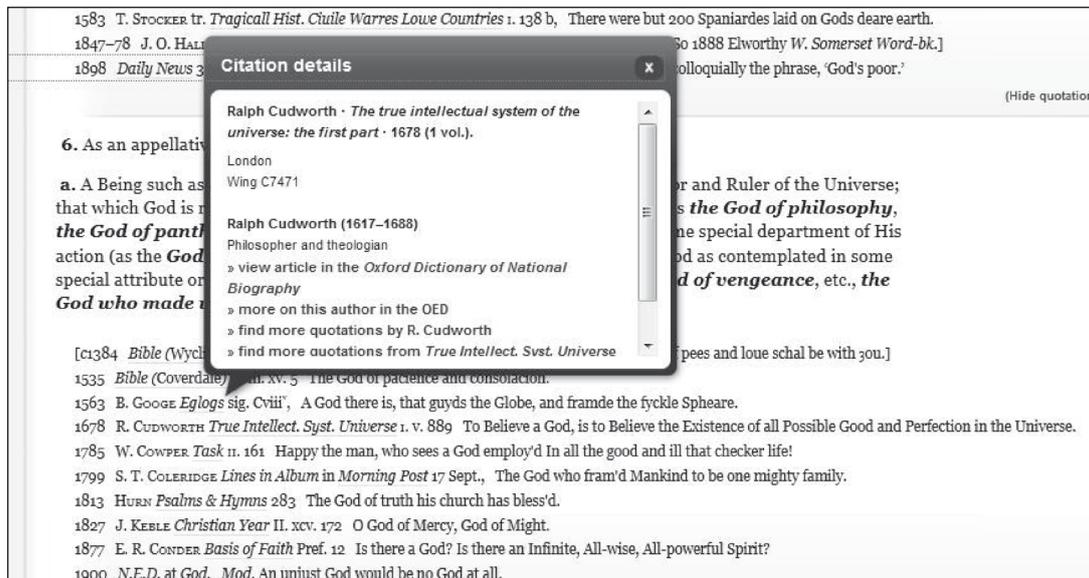


Fig. 9: Additional information on authors and works is available through a clickable link in the short title

We can end this brief survey of lexicographical functionality by looking at another area which has received extensive computational assistance, for the benefit of the reader.

There are now over three million illustrative quotations in the OED, cited normally by date of publication, author, work-title, and other bibliographical apparatus. Traditionally much of this information has been abbreviated, to save space in the print dictionary. Space is certainly still an issue; the modern lexicographer needs to present data in a concise and elegant manner to the reader, and not to use the amount of memory space available to allow the content to become loose. But there are other ways to expand this shortened data.

At present most bibliographical short titles in the text of dictionary entries have hyperlinks. If we look at the quotation dated 1678 at sense 6a of GOD n., we see it is from “R. Cudworth”, and in particular his “True Intell. Syst. Universe”. If we click on the

short-title a small box appears on the screen with additional information about the author and text. We learn that the author is Ralph Cudworth (1617-1688), and that his book is called “The True Intellectual System of the Universe”, published in one volume in 1678. Brief notes indicate that Cudworth was a philosopher and theologian, and we are invited to click another link to read his full biography in the *Oxford Dictionary of National Biography*. Further links allow you to access a specific page for Cudworth (also accessible through the “Sources” link of Top 1,000 authors and works cited in the dictionary) and to read that Cudworth is the 337th most frequently quoted source in the *OED*, with a total of 1,280 quotations in all. We also see that he is currently registered by the *OED* as the first recorded source for 197 words and 191 senses and compounds. It is also possible to click through to the quotations themselves. The reader should note that these numbers (and their accompanying data) may change at each three-monthly update of the *OED Online* database.

As a final comment we might note that the 1,141 quotations from Cudworth's *True Intellectual System* (89% of all Cudworth quotations in the *OED*) are by a long way the largest block of quotations taken from a single Cudworth work. We might then want to consider what this means: whether Cudworth is a particularly significant writer in terms of the language he uses, or whether his work was read by one of the dictionary's more dogged readers, or whether there are some other factors involved.

2. Accessing the dictionary in the future

So far we have been looking at new ways of accessing data within the dictionary and linking to external sources. We can now take a step back and examine broader issues relating to access in the future.

2.1 Access on new devices

When the *OED* was first released in machine-readable form in the early 1990s it was on the new medium of CD-ROM. This instantly brought a new tier of users to the dictionary and provided a means of advanced access to the data of a kind that simply wasn't possible previously. In the mid 1990s we experimented with options for establishing a searchable web site for the *OED* – in the early days of web-searchable databases – and finally went live with this in the year 2000.

Since then the readership – or usership – of the dictionary has broadened considerably, and we are thinking about how best to provide information for these new constituencies without prejudicing the quality of the content.

These are early days. The *OED* can of course be accessed on tablet computers, and it reads well on, for example, an iPad. The structure of the dictionary translates well to electronic devices. At present a mobile version of the *OED* is not available, and it is uncertain whether it would be satisfactory to try to represent the full content of the dictionary on a small mobile screen. If this proves not to be an option, then there is the possibility to presenting reduced “views” of the data in this format, with the full picture only available on larger devices.

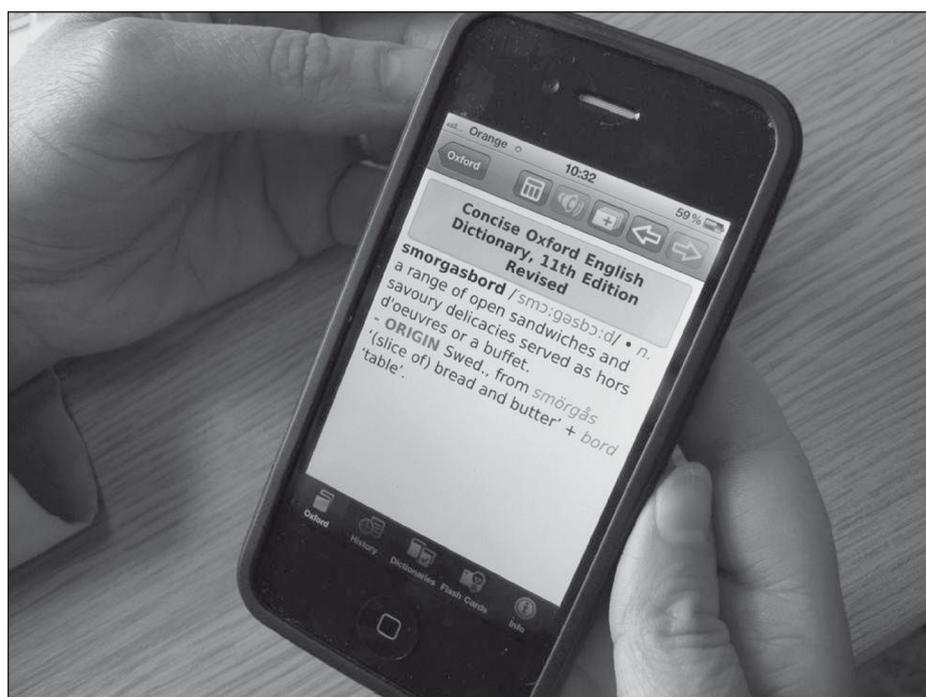


Fig. 10: The *Concise Oxford* accessible on a smartphone

But how would one go about selecting the best way to reduce the data? This is a problem that anyone working in this environment is encountering, and at present the answer is not clear as regards the *OED*. Maybe we shouldn't try to solve the problem for this generation of mobile devices, but await new technological developments which may be more satisfactory for complex language data. The present generation works well with smaller dictionaries, such as the *Concise Oxford* (see illustration).

2.2 Crowd-sourcing

Another of the ways in which the relationship between the dictionary and its readers is changing is through the medium of crowd-sourcing, or opening up an issue to as many people as possible through the Internet in order to increase the likelihood of finding a successful solution. But crowd-sourcing is not new to the *OED*. Long before the term “crowd-sourcing” was invented, the *OED* was appealing to English speakers throughout the world to contribute to the compilation of the dictionary, especially by providing the editors with additional evidence of language use from historical texts.

This procedure has proved extremely successful over the years, and the *OED* has benefited immeasurably for the assistance of the general public. The new media allow us to extend this idea, and for several years now science-fiction aficionados have had their own wiki through which they can publish their lexical findings and be sure that these are considered for the rolling revision of the dictionary.

At the *OED*'s sci-fi web site (www.jessesword.com/sf/) there are pages for each of the *OED*'s science-fiction terms, along with an explanation of what new material would be helpful. There are also pages for terms not yet in the dictionary but for which readers might like to contribute information.

SCIENCE FICTION CITATIONS

[\[Home\]](#) [\[How to Cite\]](#) [\[Newest\]](#) [\[Date Graph\]](#) [\[About\]](#)
[\[Science Fiction\]](#) [\[SF Criticism\]](#) [\[SF Fandom\]](#)

Full record for pressure-suit n.

Definition	=space suit
OED requirements	antedating 1928
Earliest cite	E. E. Smith <i>The Skylark of Space</i>
Comment	Alistair Durie submitted a 1928 cite from E.E. Smith's "The Skylark of Space". Earliest cite in the OED is from 1936
Last modified	6 July, 2008

Citations for pressure-suit n.

click [here](#) for more information about the citation list

1928 <i>Amazing Stories</i> 3 539/2	Have you fur pressure-suits?
1969 R. A. HEINLEIN <i>Let. 12 Nov.</i> in R. A. Heinlein & V. Heinlein <i>Grumbles from Grave</i> (1990) 196	I lured him... by promising him that he could help develop pressure suits for fighter planes... It led directly to him developing the first suit used on the Moon.
1969 R. A. HEINLEIN <i>Let. 12 Nov.</i> in R. A. Heinlein & V. Heinlein <i>Grumbles from Grave</i> (1990) 196	I lured him... by promising him that he could help develop pressure suits for fighter planes... It led directly to him developing the first suit used on the Moon.
1976 C. HOLLAND <i>Floating Worlds</i> (1977) 461	A man in a pressure suit jogged across the gray dust to the ship.
1992 V. VINGE <i>Fire upon Deep</i> II. xxv. 189	The pressure suits he built had power packs and weapons stores.
1992 B. BOVA <i>Mars</i> (1993) 3	Through the thick insulation of his pressure suit, Jamie could hear nothing except his own excited breathing.
1994 B. BOVA <i>Craft of Writing Sci. Fict. that Sells</i> IV. 22	With a twist of his shoulders to settle the weight of the pressure suit's bulky backpack, he shambled over to the packet of seismic instruments.
2004 P. F. HAMILTON <i>Pandora's Star</i> IV. 93	The last half had to be covered in pressure suits; and the only real view, sensational though it was, came from Aphrodite's Seat, the clifftops just below the caldera plateau.

Fig. 11: Evidence for the expression *pressure suit* collected for the *OED*'s science-fiction wiki site

We can take as an example the expression *pressure suit* (defined by the *OED* as “an all-body garment that can be made airtight and inflated to protect the wearer against low ambient pressure (as in high-altitude flight)”). The previous printed version of the *OED* (1989) offered an earliest use for the term from *Flight* magazine in 1936. Readers have contributed numerous further examples of the expression, the earliest of which from 1928.

In future it would be possible to extent the concept of dedicated (and controlled) subject-specific wikis to other areas of language for which there is an active community of users eager to contribute to the *OED*, or simply to publish their findings for others to see.

The earliest appeals to the public from the *OED* in the late 1850s requested volunteers to read through specific texts extracting data that would be useful to a dictionary intended to describe the English language throughout the world and over many centuries. But soon these appeals turned into lists of desiderata, printed and circulated in the hope that members of the public would know just that piece of information that the editors required.

More recently, appeals have also been made through other media, the most extensive being the BBC2 television series *Balderdash & Piffle*, aired in 2006 and 2007. In this series a set of words was issued as a challenge to the public, who were asked to provide printed or other verifiable earlier evidence for the existence of the terms. The show was very successful, with many of the appeals (such as *ploughman's lunch*, *something for the weekend*, and *made-up* (= “happy, pleased”)) proving successful. In 2012 a new on-going Appeal was launched. After starting with a handful of active appeals, new requests are regularly added (*blue-arsed fly*, *party animal*, *blues and twos*, etc.) and findings are fed back to the editor for incorporation in the dictionary.

OED **appeals** Help record the history of our language

[Home](#) → [Appeals](#) → [easy-peasy](#) ↓

easy-peasy adjective earlier than 1976 [View OED entry](#)

This colloquialism for 'extremely easy' was added to the dictionary in 2002. *OED* editors posted an appeal for earlier evidence in our print newsletter back then, and we haven't given up. We're still trying to find evidence earlier than our current first quotation:

1976 *Evening Standard* 20 July 19 (caption) I'm short of money this week. How do I get to work? Easy peasy. I run to the station, leap the barrier, clobber the train driver, dress up in his clothes [etc.].

Earlier evidence may be available in print, or in new digital archives which were not available when we initially researched the entry.

Posted by OED_Editor on 28 November 2012 10.12
Comments: 3

Fig. 12: An online appeal launched by the *OED* in November 2012

In the light of the success of both these models, it is possible to imagine in future that the dictionary will operate some form of wiki-based system for collecting information from the public about any English word or meaning, while at the same time running directed appeals for specific words for which editors actively need more data.

This in turn ties in with another idea which is in general circulation in the reference community, and that is the publication of so-called “stub” entries, as in Wikipedia. Publishers and editors are aware that the expectations of readers have changed over the years with their growing familiarity with information obtained from the Internet. Whereas previously the assumption was that published information should be fully researched and edited before it was released, there is now an expectation that although finalized, authoritative text is required it may also be acceptable to show unfinished work, as long as it is properly identified. Readers would therefore be able to see which new words, for example, might be edited in the near future (and for which they may be able to provide additional information), or they may even be offered the chance to provide a new or improved definition for a term (as they would expect to do in any case on projects such as the *Urban Dictionary*).

This is not to signal a free-for-all in publishing. Information would need to be flagged with regard to its level of authority, and all work would still be moving towards the fully finalized entry. But readers would perhaps benefit from seeing how work is progressing, and they could also contribute to moving the entry on to the next stage.

At present the *OED Online* is provided on the basis of a subscription model. In the UK and New Zealand most public libraries offer access free to their readers through separate agreements with Oxford University Press. But the expectation of many on the Internet now is that content should be provided free of charge to the end user. This appears to offer content providers little scope for directly recouping costs – which can be substantial.

Different providers are examining different ways to overcome this obstacle. At the two extremes we find funded bodies (often national language bodies or universities) posting content without charge to the user. At the other extreme, substantial charges are levied which often result in even large universities and other substantial organizations being unable to afford important resources.

There are any number of options within these two alternatives. Charges can be levied on individual searches or by results; shorter- or longer-term subscriptions can be offered; the resources can be separated into ‘premium’ content, for which charges are incurred, and regular content, which may be searched free, perhaps alongside paid advertisements.

There are major issues involving user satisfaction here. Whilst the user may appreciate being able to run basic searches on an important resource, he or she may become increasingly frustrated at being prevented from drilling down into results without further payment drip-feed. But others will only ever need to look at one type of information (perhaps definitions in the case of dictionaries), and so a publisher might be unwilling to satisfy 100% of the needs of this class of user without any charge whatsoever.

The *OED* is still considering all of these options, as are many other data providers. We would expect to emerge from this debate with new plans, but at present it is uncertain what these would be.

3. Conclusion

The *OED Online* is a major resource for the documentation and study of the English language. As a result of its most recent functional update it offers a broad range of search strategies (often based around the notion of filtering down from previous results). It also offers much additional discursive information on the dictionary and the language, and digests of material analysed from the main dictionary database. Over the years we can expect to see the diversity of this functionality and information extended.

Many of the issues addressed and being addressed by the *OED* are common to other national historical dictionaries and, indeed, to any organization offering material on the Internet. As well as enhancing the dictionary content by regular quarterly updates we can expect further software development to expand the ways in which data can be accessed or displayed on the site. Furthermore, we might reasonably expect further links to other historical dictionary sites outside the English-speaking world, and to other types of sites (holding modern or historical text, encyclopaedic or other reference information, images, etc.).

As well as this, data providers need to examine more closely the options for making proprietary data available over the Internet, whether this is by fully open access, free ac-

cess to restricted segments (however generated), premium charges for specialist data, or by some other mechanism. This is not the sort of issue which has only one answer, and we can expect to see different answers testing the water. Perhaps the final outcome will be influenced by the fact that although we are data providers, at heart and at bottom we are all users ourselves.

