Philos. Technol. (2013) 26:341–348
DOI 10.1007/s13347-013-0118-7

COMMENTARY

The *Oxford English Dictionary* and Its Philosophy

John Simpson

Received: 24 June 2013 / Accepted: 25 June 2013 / Published online: 7 July 2013
© Springer Science+Business Media Dordrecht 2013

Keywords *Oxford English Dictionary* · Philosophy · Principles

1 Introduction

Does the *Oxford English Dictionary* (*OED*) have a "philosophy", in the sense of a set of guiding principles? The original editors of the *OED*, in the latter part of the nineteenth century, would have pointed in explanation to the full title of the dictionary:

> *A New English Dictionary on Historical Principles; founded mainly on the materials collected by the Philological Society.*[1]

The keywords are "on Historical Principles", an expression over which there was much discussion before the first page of the first instalment of the dictionary was published in 1884. The *OED* wished to distinguish itself from contemporary dictionaries ("synchronic" dictionaries, concentrating on one period of the language). The *OED* was a "diachronic" dictionary, addressing language across its history. This historical approach meant that the editors took note not only of how the language stood in the present. Every aspect of the dictionary's content was informed by a historical perspective, from the anterior etymology through to the latest illustrative quotation.

[1]The *New English Dictionary* (*NED*) was officially retitled the *Oxford English Dictionary* (*OED*) in 1933. The dictionary was edited over 44 years (1884-1928), and a one-volume supplement was added to the "corrected re-issue" of 1933, when the title was altered. The words "Oxford English Dictionary" had however appeared in a subsidiary position of the title page of instalments of the dictionary from 1895.

Reference to "the Philological Society" reminds us that the dictionary grew out of two lectures delivered by Richard Chenevix Trench (later Archbishop of Dublin) to the Philological Society in London in 1857. The Society was tirelessly engaged in collecting materials for the dictionary and its involvement continued unofficially for several years after the contract to produce the dictionary was eventually signed by the Oxford University Press in 1879.

J. Simpson (✉)
Oxford English Dictionary, Oxford, UK
e-mail: john.simpson@oup.com

This was an understandable perspective for the reader, but what does it imply for the lexicographer, the editor? What problems does it throw up? And is it a valid approach today?

2 How Does a Historical Dictionary Map the Language?

A synchronic dictionary presents the reader with a snapshot of a large number of related lexical strands at one point in time (typically the present), but it is not usually possible to see how the strands are related. This does not matter to the user of the dictionary, who is principally interested in how a word is used in the time period covered by the dictionary.

A historical dictionary can, however, contain a more extensive view, based on an assumption that is applied at every stage of the editorial process: that there is an incremental and logical progression (of meaning, spelling, pronunciation, etc.) throughout the chronological history of any word, from the earliest times right up to the present day. Some aspects of this progression may play out in the language from which English "borrowed" a word (and English may then appear to borrow senses almost at random). The historical lexicographer's task is to describe each word whilst bearing in mind this concept of incremental logical progression over history.

Within the history of lexicography this view came to prominence at the same time that numerous other disciplines were adopting a similar chronological/logical imperative. The emergence of "comparative philology" in the early to mid nineteenth century—examining how changes in one language are paralleled by changes in others, and pointing to interrelationships and a historical progression—found expression at the time in archaeology, geology and other disciplines, and is perhaps best recognized in Darwinian evolution.

In the twenty-first century yet another concept is beginning to be realized for the dictionary, as online it takes its place as one of the hubs of information on the Internet across which users may follow links between different types of text (literary or newspaper text, reference data, images, etc.). Such a position for a historical dictionary could hardly have been foreseen when work began on the first edition of the *OED* well over a hundred years ago.

3 The Historical Principle in the Entry for Culture

The following section summarizes briefly how this concept of incremental logical progression over time works for the English word *culture*. *Culture* entered English, according to the records, in the Late Medieval Period (the first reference in English is dated *c.* 1450). It came both directly from Anglo-Norman and Middle French *culture*, and also from classical Latin *cultūra* (from which French itself derived), or (and this is perhaps more likely) from a mixture of both sources over time.

Are we surprised that *culture* is not recorded in English before the middle of the fifteenth century (i.e. can we assume this is a logical time for the word to arise?)? A Latin/French (Romance) word is very unlikely to have been used in English in the Old English (Anglo-Saxon) period, before the Norman Conquest, when the vocabulary of English was largely Germanic. Arguably *culture* might have occurred before

1450 (and indeed it doubtless existed outside the surviving written record), but perhaps the peasant occupation of cultivating the soil did not attract a Romance description till late in the medieval period. Until then, the concept of "cultivation" was well covered by earlier Germanic-based words such as *tilth* and *earth-tilth* (Old English), *tilling* (?*c.* 1225→) and *land-tilling* (*c.* 1420), and *delving* (1377).[2]

Lexicographers are trained to look for specific features at the point of transmission of a word from one language to another. According to the dictionary's philosophy we should expect (at least approximate) identity of form and pronunciation at the cross-over point; that the borrowed meaning should pre-exist in the donor language; some geographical or cultural reason for the transmission; and an explanation if any of these and other characteristics are absent. A review of the documentary evidence and social context of the period shows that these conditions are met in the case of *culture*.

There is very unlikely to be a single point of transmission. Word borrowing occurs over time, and sometimes over centuries, as new meanings are borrowed at different times. The *OED*'s philosophy of language accommodates long-term inter-relationships between languages (and their accompanying cultures) rather than single explosive interaction.

After amassing and sorting the available materials for the word *culture*, it is possible to determine that the semantic development moves (as can be expected) in short jumps. The basic meaning of the word in English (and this meaning was current in the donor languages) concerns the cultivation of the soil. By a small semantic shift, we have a new sense 100 years later (1580 or so): the cultivation of crops. By another shift, we have the artificial propagation of microorganisms (1880 onwards). This last shift coincides in time with the emergence of scientific research into microorganisms (the word also dates from 1880). Social and cultural change and the language used to describe it often walk hand in hand.

If we look at another branch of the "genealogical tree" of *culture*, we see that a minor leap from "cultivation of the soil" (late medieval) gives us a new strand of meaning: the cultivation of the mind (early sixteenth century) and later the refinement of the mind. Each time a single attribute or feature of the word changes. Note that it is (theoretically) a necessary condition that the base meaning predates an extended meaning, as is the case here and as is normally satisfied by historical documentary evidence. If this evidence is lacking, then the lexicographer needs to examine the reasons for the disjunction. The reasons may be evidentiary or cultural/social.

Once *culture* has leapt away from agriculture and husbandry, it is on the open road for development in the "social custom" sense we are very familiar with today. Curiously, this development, in the mid-nineteenth century, involves the influence of German *Kultur*. The word had undergone its own semantic changes in German over the previous hundred years, and their reborrowing into English helps to explain a semantic change which would otherwise be hard to explain.

This form of progression is apparent in almost any word in the language, and the dictionary's philosophy has been developed to accommodate it.

[2] See the *Historical Thesaurus of the OED*, which forms part of the *OED Online*.

4 Computers and Language

Are computational linguistics and online editing compatible with the *OED*'s philosophy of language?

Traditionally, material was collected for historical dictionaries by means of reading programmes, whereby text extracts from historical and modern sources were transcribed on to index cards and then filed centrally.

Today's *OED* maintains a hybrid system in which a series of card files (much less dominant than in previous decades) are used in parallel with enormous online databases (corpora) of historical and modern text, searchable by increasingly "intelligent" software tools.

The availability of these databases allows the editors to know that the data they analyse are more complete than the data available to earlier generations of *OED* editors. This permits a more fine-grained semantic progression to be documented. Arguably the problem has shifted from the relative paucity of data to its dramatic overabundance. In both cases, editors have needed to develop practical methods to filter and sift the material.

The *OED* itself was conceived as a book, published in a series of volumes. This was the case from the first edition of the late nineteenth century (1884 onwards) until the second edition of 1989. After that, the dictionary became available first on CD-ROM (and magnetic tape) and then online.

The fact that the dictionary could transfer relatively easily from a traditional to a computer-based medium is largely due to the design of the printed version. In comparison with other historical dictionaries started earlier, the print *OED* contains data organized in what can easily be regarded as data fields, and in a structure that transfers conceptually to a fielded, structured computer database.

At the present moment of its development, the online *OED* is moving away from appearing to be a book that has been placed on the Internet, to becoming a purely computational resource. It has not yet entirely shed some of its book-like features, but as time goes on this will doubtless happen (even though one "view" of the data may remain a book-like view).

Scripts written to assist editorial work on the revision and update of the dictionary support the philosophy of incremental logical progression over time outlined above. Computational tools are easily able to determine whether meanings have been placed in a chronological tree structure. If this rule is breached, the editor has to review and restructure the data, often as a result of further research. New nodes are simply created within an entry to accept newly edited meanings, compounds, etc. And the overall structure allows completely new entries to fit snugly into the existing dictionary.

When additional new or supplementary material is uncovered once an entry has been revised and published, the conceptual dictionary supports its inclusion, and various editorial and computational checks are made to ensure that the structure is not breached.

5 Illustrative Quotations and the Dangers of Inexactitude

Various factors can derail the accurate historical analysis of language. This section looks at one such factor.

The majority of the text in a historical dictionary such as the *OED* consists not of definitions or etymologies, but of historical and modern illustrative quotations documenting the use of senses of words over time, from their earliest recorded use. It is these quotations which form the backbone of the dictionary text and the raw material from which any analysis of the language arises. Lexicographical decisions should typically be based ("descriptively") on the evidence of the language as displayed in the raw data (the illustrative quotations).

Here is one of the illustrative quotations accompanying the earliest (and now obsolete) sense of *premeditation* ("the action of thinking about or contemplating something beforehand or previously, without implication of purpose"). The sense is obsolete because it carries no implication of purpose, as the word does today.

> 1685 C. COTTON tr. Montaigne *Ess.* (1877) I. 82 The premeditation of death is the premeditation of liberty; he who has learned to die has unlearned to serve.

The First Edition of the *OED* presented this seventeenth-century quotation through the eyes of a Victorian editor in an edition dated 1877. The editor has modified the text to make it easier to understand for his nineteenth-century readership. But any modification of original text can invalidate later lexicographical analysis.

After conversion to the first edition, this quotation alters to the following (changes marked in bold):

> 1685 C. COTTON tr. Montaigne Ess. **i. xix. 113** The **P**remeditation of **D**eath is the **P**remeditation of **L**iberty; who has learnt to d**y**e has **forgot** to serve.

These changes are certainly significant, though they do not in this case affect the dictionary's treatment of the headword. The syntax has changed ("he who" to just "who"); a complete word has changed ("unlearned" corrected to "forgot"); and various other spelling and capitalization changes are made. The change from "forgot" to "unlearned" is Victorian, but eighteenth-century authors took other liberties with the quotation, preferring "forgot to be a slave" to "forgot to serve", though the Victorians changed this back.

Very often later modifications alter the text substantively. The historical lexicographer should always guard against mediation and modification of text.

6 What Is Absolute Data?

Readers make the reasonable assumption that dictionaries are authoritative. Dictionary editors like their data to be as accurate as possible. But almost all editorial text in a dictionary is relative, not absolute.

Which sections of a dictionary constitute absolute data—data which are absolutely accurate over time, and not simply accurate at the time the dictionary is compiled?

The definition is a central part of a dictionary entry, and yet different dictionaries define precisely the same term in different ways. So there is clearly not one single "definition" that is right for any given word or meaning. In addition, different definitions may be used for different levels of user (e.g. learner's dictionary v. school dictionary).

Furthermore, it is possible for definitions to be split into sub-definitions, or for a number of separate meanings to be coalesced under a single heading. From a

historical perspective, meanings (and the contexts in which meanings exist, and our understanding of a term) change gradually over time: sometimes our interpretation of the definition keeps pace with this and at others some editorial intervention is required to close the differences.

In addition, the lexicographer knows that many illustrative quotations are informed by more than one meaning, and so they become poor examples for the purpose of exemplification. There are gaps between definitions, into which some illustrative quotations fall. Language does not consist of a series of distinct meanings, but of a semantic spectrum of meaning many examples of which cluster around particular dictionary meanings. In order to stave off lexical chaos, it is helpful to imagine a fixed set of defined meanings (as a dictionary does), and a series of sub-meanings which may well continue to represent nuances and variation below the radar of account in a formal dictionary.

Are illustrative quotations absolute data, once they have been converted to their first edition? The text published in a first edition is not the original text. Should editors of historical dictionaries make a point of returning to holograph manuscripts (if they exist), where text may be slightly different from that printed. There may be several manuscripts representing different stages of the same text, not all of them preceding the first printed edition. Should the "authoritative" date given for the quotation be the date of publication or the date at which a text was first written?

Another conundrum for editors is that a text may predate the appearance of a given word in that text; a performance of a dramatic text may predate the definite occurrence of a given word in the printed text of the play. The lexicographer is hardly concerned here with the larger picture (the original or performance of a text), but only with the appearance of a lexical item in that text. This can lead to different ideas of the appropriate date by which a quotation should be labelled. The *OED* takes a conservative view and the data remain relative.

Pronunciations are not absolute over time. Written language is itself an artificial construct, self-regulating (as regards English at least) rather than subject to an equally artificial academy. In fact, the closer one looks at every aspect of an authoritative historical dictionary (or any other dictionary—and perhaps any other work of reference) the more one realizes that nothing is absolute and everything is subject to change. In terms of the lexicographer's view of the data, a dictionary presents a transient coherent structure composed of a series of pieces of relative data.

The upshot of this is not that dictionaries cannot and should not be trusted, but that the nature of language and of our interpretation of history is such that what appears right, on the best evidence, today may well represent an outmoded view in years to come.

7 Where Are the Boundaries?

Is the *OED* based on raw data collected only from printed texts, even if we include printed editions of text originally read in manuscript before the dawn of printing? Clearly not: the historical lexicographer will collect data from wherever it is available and verifiable (and datable, preferably).

Printed text has long been the resource of choice for historical lexicographers, if only because it is easily reverifiable by others who wish to replicate the lexical

analysis. But whereas we used to think of libraries as the principal archives, text (in whatever form) is now often securely archived electronically. This makes it easier for the historical lexicographer to cite different types of data in a way that can still be subject to public, scientific scrutiny.

Several years ago *OED* guidelines changed, allowing editors to cite reliably dated text from the Internet in dictionary entries, especially when this represented the earliest record available for a word or sense. With these boundaries relaxed, it became possible to cite this reference, for example, as the earliest known occurrence of the new term "blogosphere":

> 1999 B. L. Graham www.bradlands.com (Weblog) 10 Sept. (O.E.D. Archive), Goodbye, cyberspace! Hello, blogiverse! Blogosphere? Blogmos?

We would be corrupting the record not to provide this first quotation.

In the same way, data can come from other irregular sources, such as historical photographs published in a later book. The *OED*'s first reference to "hot to trot" comes from a sign on a photograph dated 1950 in a text published in 1993:

> 1950 in R. Tomedi *No Bugles, No Drums* (1993) xix. Plate 2 (*photograph of U.S. Army company sign*) Hot to trot 2d B. 7th Reg. Fox co.

When there are reliable archives of the spoken word, then the *OED* is happy to cite oral references.

8 The Lexicographer's Critical View

Everything leads to the requirement that the lexicographer remain critical of every piece of data that is encountered. Both kind-hearted and aggressive critics of historical dictionaries often misinterpret text, meaning that their suggestions or criticisms are often (though clearly not always) based on shaky observation.

The lexicographer needs to maintain a rigorous scepticism and neutrality about everything, from the possible (multi-)interpretation of a snippet of obscure fifteenth-century text to the apparently simple pronunciation of a contemporary monosyllabic word. Nothing can be accepted until it proves itself.

The structure and content of a dictionary entry theoretically derive from the information held in the raw data about the language collected in the form of quotations. Every decision should (it is maintained) arise from an interpretation of the quotation material. But even this is open to discussion, as the lexicographer's world knowledge and other external data have to inform each definition. It is impossible to define a rose simply from a collection of illustrative quotations. The information garnered from illustrative quotations is central to the lexicographical interpretation of a term, but it is only part of the real-world material needed to interpret reality in a dictionary. The lexicographer needs to determine which attributes of a term are required within the definition, and which are probably incidental or extraneous. The balance will differ depending on the dictionary, the intended audience or any of a number of other features. The illustrative quotations may contain skewed data, giving more relative weight to one factor than the lexicographer considers it should have. They may mix definitional and (wider) encyclopaedic

information, which the lexicographer may want to separate. They may on rare occasions even mislead. Every entry is a problem, which must be resolved with whatever valid information is available.

9 Conclusion

The overall question "Does the *OED* have a philosophy" merits investigation because it is too easily ignored. The mass of detail means that the philosophy can be overlooked. The *OED*'s policy was established to handle a language that evolves by logical progression over time. The resultant data has proved useful to scholars and other readers, which is perhaps one measure of success. But most readers remain generally unaware of how a historical dictionary works (sometimes lamenting the absence of ahistorical features, such as the parading of the most frequently used meanings, for example, first within each entry).

This is to misunderstand gravely the way in which a historical dictionary, or at least one which is able to cover the full chronological spread of a language, can operate. We must assume that B cannot exist before A has evolved; we must accept that one thread of meaning may bifurcate at any moment, but that there is usually an internal or external reason for this change. When there is a disjunction between the evidence and the logic, then this is likely to represent a failure of evidence or logic—or is perhaps a suggestion that the lexicographer is dealing with a type of new structure. But these new structures occur very rarely.

Lexicographers adopt various methodological strategies to ensure that the conditions for lexical analysis are optimal: the raw data should show as little editorial intervention and interpretation as possible; a conservative approach is typically regarded as the safest route; but the output of lexicographical work is temporary—ideally the best that can be achieved at the time of editing and with the materials available.

Most readers of a historical dictionary are not aware that their dictionary has a philosophical dimension, though they may be aware that somewhere along the line a complex methodology has been invoked. For many readers this does not matter. They find the information they need and move on to other concerns. But for a small percentage of readers interested in the interaction of their language and their culture, the structure and intention of their dictionary is of vital importance and helps to inform a more intelligent view of what can seem chaotic data.