

Johnson, the Dictionary, the Wiki, and the Web

Samuel Johnson would have loved the internet. Apart from anything else, he'd have loved the scale of the thing, its sheer size. It's the largest collection of linguistic material there's ever been. There's more data on the internet than in all the libraries of the world combined. Johnson would have loved to bathe in its diverse content. He'd have wanted to sort it out. The size of the task wouldn't have put him off. As he said in the Preface to his Dictionary, 'A large work is difficult because it is large'. You can almost see him salivating. And certainly, nobody had ever attempted such a large-scale lexicographical enterprise before. Over 42,000 entries in its first edition. Over 140,000 definitions. Over 220,000 illustrative quotations. It set new standards for dictionary writing.

Johnson's project has been called the first modern dictionary. But what does that mean, exactly? What is it that made it modern?

There are two sides to dictionary compilation: coverage and treatment. Coverage refers to the number and range of words explained in a dictionary. The blurb will tell you: this dictionary contains 100,000 words, or 200,000 words, or whatever. And the title will tell you: this is a dictionary of legal English, or a dictionary of spoken English. Treatment refers to the *kind* of information included about those words - their meanings, pronunciations, etymologies, stylistic nuances, and so on. Johnson transformed both of these areas. He dramatically extended the coverage of a dictionary, and he absolutely revolutionized the way entries were treated.

What was dramatic about his coverage? He did something so obvious (to modern eyes) that it's hard to believe it hadn't been done before. He included the 'easy' words - words that everyone knows, like *what* and *of* and *three* and *January*. Words that play an important part in English idiom, such as *make* and *do*. Dictionaries had simply ignored these words before, concentrating on the 'hard' words that people might *not* know, such as *allegation* and *circumlocution*. Johnson wanted complete coverage. And, of course, it turns out that the easy words aren't easy to deal with at all. In fact they're the most difficult ones to describe, because they have so many uses. Take the verb *take*. This is the longest entry in his Dictionary. It takes up 11 columns of print, and has an amazing 134 uses, including many phrases using this verb - the first time these see the light of day in a dictionary:

to take after - to learn of, resemble, or imitate

to take care - to be careful

to take in - to comprise or comprehend

to take in hand - to undertake

to take off - to invalidate, destroy, or remove

to take up with - to lodge or dwell [FADE]

Such mammoth entries were unprecedented in English dictionaries, and they're remarkable in the way they describe the subtle nuances of meaning conveyed by these everyday words.

But Johnson did something more. He showed how people actually used the words. He and his small team of assistants spent thousands of hours searching through the works of major authors, finding examples of every word, and painstakingly transcribing them onto slips. For example, he found *scarecrow* in Shakespeare:

We must not make a scarecrow of the law, / Setting it up to fear the birds of prey

and also in Dryden, Spenser and Raleigh too. Today, this is routine lexicography. Dictionary writers compile a collection or *corpus* of examples, and use these to make their entries come alive. It's Johnson's method exactly - except that modern dictionaries pay proper attention to the spoken language. In Johnson's time, only written English was thought to be worthy of study. But the principle is the same, that dictionary-writers shouldn't rely on their own heads to cook up artificial examples, but should look at the way words are actually used. It's meticulous, time-consuming work - these days, of course, made so much easier by the ability to search for words and phrases on the internet. Imagine it all being done with ink and quill pens. You'll probably remember that Johnson himself described lexicographers as 'harmless drudges'.

In fact, lexicography is far more than drudgery. It's a highly creative enterprise. It's a huge intellectual challenge to capture the meaning of a word precisely and to express that meaning clearly. Johnson was brilliant at doing this, though he was in no doubt about the challenge. *Finding* the words is relatively easy, he says in his Plan for the dictionary:

The great labour is yet to come, the labour of *interpreting* these words and phrases with brevity, fulness and perspicuity.

It was indeed a huge labour, and when we look at a set of Johnson's definitions today, it's obvious how much thought must have gone into them. They are the dictionary's primary strength, and its chief claim to fame. We can easily get a sense of the problem by trying to formulate for ourselves appropriate definitions for such words as *nature*. Think about it. How on earth does one begin to define a word like *nature*? Abstract words pose particular problems, but all words require definitions that are clear and succinct, and Johnson's achievement can be seen on virtually any page. Here's his definition of *message*:

An errand; any thing committed to another to be told to a third.

Or his elegant definition of *history*:

A narration of events and facts delivered with dignity

His definitions can also be humorous. Listen to the cheeky alliteration when he defines *heresiarch*:

A leader in heresy; the head of a herd of hereticks.

There's another side to Johnson's dictionary that makes it feel modern. He saw the importance of regional and social variation. He included several regional dialect words, including some from his home county of Staffordshire, such as *moreland*, for a type of hilly countryside. And he gives a great deal of information about the social variation of his time - observations about eighteenth-century usage - or, at least, Johnson's opinion about contemporary usage. The stylistic range of the *Dictionary* is in fact very wide. At one extreme we find highly formal words of classical origin, such as *adumbrate* and *prognostication*; at the other we find colloquial interjections, such as *hist* and *tush*. He even notices some gender differences - *frightfully* and *horrid*, he says, are 'women's words'. At the same time, being part of the spirit of his age, he routinely draws attention to words he considers 'low', 'vulgar', not to be used by well-brought-up people in 18th-century society - words such as *budge* and *desperate*. These words have lost their vulgar associations today, so it's really interesting to learn that they were socially sensitive words then.

Johnson gave regional, social, and stylistic variation a presence in his dictionary that hadn't been seen before. And he knew what he was doing. In Chapter 20 of James Boswell's *Life of Johnson*, we find this exchange between the two:

[Johnson:] By collecting those words of your country, you will do a useful thing towards the history of the language.

[Boswell:] He bade me also go on with collections which I was making upon the antiquities of Scotland.

[Johnson]: Make a large book - a folio.

[Boswell]: But of what use will it be, sir?

[Johnson]: Never mind the use; do it.

And it is this which makes me think that Johnson would have felt very much in sympathy with the linguistic dimension of the internet, for it is giving a home to regional and social variation in English in an unprecedented way. Think back a decade: if you wanted to find out about any regional dialect, or to hear examples of a regional accent, how would you have set about it? It wouldn't have been easy. Now you can find dozens of sites for all the major dialects, most of which allow you to download examples of the local speech. And by 'major dialects' here, I mean two things: major national dialects, such as Scots, Yorkshire, and Geordie; and major international dialects, such as Australian, Indian, and Singaporean English. We mustn't forget that, of the 2 billion users of English in the world, three-quarters are not native speakers, and the distinctive features of their emerging dialects is just as much a part of the 'English language mix' as are the older distinctions such as British vs American. All of course are now easily accessible via the Web, and colloquial forms of these dialects can be encountered in a multitude of chatrooms and social networking sites.

Johnson would have been both delighted and horrified by the internet. He would have loved its diversity, not only within a language but as a repository for all languages, including those that are seriously endangered. Do you remember this famous quotation?

I am always sorry when any language is lost because languages are the pedigree of nations.

On the internet, of course, nothing gets lost. It's the ideal medium if you want your language, no matter how small and unimportant it might be, to receive a global presence. And increasingly, minority languages are finding their way onto the internet. Johnson would have highly approved of that.

But I think he would have been highly disapproving of another internet phenomenon: the user-generated dictionary - the 'wiki'-dictionary. This is where you and I can call up a dictionary site and add a word to it, or say how we use a word. We can cook up our own definitions and put them online. Now these can be very useful hints about the way the language is being used, but usually the information is of little value. As I've said, it isn't easy to phrase a definition clearly and succinctly, and most of the definitions I've seen lack the kind of rigour we associate with Johnson. The internet can learn something from Johnson in that respect.

But there's something which would have upset him even more. In order to fully appreciate the use of a word you have to know who is using it. Is it a young person's word? Is it a word used by men or women? Which part of the world does the user come from - think of the differences between British and American English, for example, to see the importance of that. There are lots of variables - and the problem is that the internet doesn't give you the data, mainly because most contributions are anonymous. The internet revels in anonymity. We usually have no idea whether a post arriving at a dictionary site is coming from a London teenager or a San Francisco grandmother, or whether the sender is a native speaker of English at all. Johnson was scrupulous about identifying the sources of his quotations. The internet can learn something from him in that respect too.

So, to my mind, people are quite wrong to think of Johnson's Dictionary as a 250-year-old antique, useful only as a guide to the 18th-century, or important because of its historical influence, and no more. Johnson's principles and procedures are highly relevant to the way we think of words today. And most of all, he tells us to remember his own experience in dictionary-writing, and to develop a realistic sense of what dictionaries are all about. He learned this the hard way. At the beginning of his project, he had been a purist. In his Plan for the Dictionary he's absolutely positive:

one great end of this undertaking is to fix the English language.

To fix it. To stop it changing. It's an attitude still widely encountered today. But by the time he'd finished the Dictionary he'd changed his mind, and in his Preface he acknowledges how absurd this notion had been:

With this consequence I will confess that I flattered myself for a while; but now begin to fear that I have indulged expectation which neither reason nor experience can justify.

Language changes because society changes. The only languages which don't change are dead ones. It's the job of dictionaries to record these changes, accurately and honestly. And if there's one message which Johnson wants us to remember, over the centuries, it is, as the modern idiom puts it, to 'get real'.