

'Oh, yeah, stupid': Bloggers help to update Oxford Dictionaries

In order to **really** understand how our language is developing during the 21st century, researchers at Oxford Dictionaries are monitoring the whole of the English language as it is being used by everyone, everywhere, every day. They're looking not only at newspapers, magazines and fiction but also blogs...

Since 1 January 2000 the researchers have fed over 1.8 billion words of what people around the world are writing and saying into the Oxford English Corpus, a unique electronic database that makes it possible to see exactly how and why English is changing.

The importance of blogs

What do blogs tell us about how people are using language? Is the language of blogs different? One way to find out is by analysing and comparing blogs with other sorts of writing. With 1.8 billion words of English, and over 175 million words from blogs in the Oxford English Corpus, we can start to see what patterns are emerging.

Recently, researchers from Oxford Dictionaries have been studying the language of blogs in the Oxford English Corpus, and here are some of their findings:

Me, myself, I

It may come as no surprise to find that bloggers are, well, interested in themselves. The pronoun 'me' is *five times* more likely to be used in blogs than in other sorts of writing, while 'myself' and 'I' are both significantly more used by bloggers. And, as bloggers, we are on the whole more interested in expressing our opinions than in talking about facts: adjectives such as 'stupid', 'lovely', 'nice', 'interesting', 'odd', and 'wonderful' – as well as words like 'suppose' – are staple words in a blogger's vocabulary.

The write 'stuff'

'Stuff' is the sort of vague word that many writers love to hate, but bloggers, well, just love it. 'Stuff' is more than *five times* as common in blogs as in other writing. And bloggers routinely use adverbs such as 'somewhere' and 'somehow'; it seems that to blog is to avoid being precise.

Bloggers' vocabulary

The top 15 words you're more likely to use in blogs, compared to other forms of writing...

1. blogger
2. blog
3. shit
4. oh
5. yeah
6. stupid
7. post
8. ok
9. stuff
10. lovely
11. myself
12. update

13. nice
14. me
15. my

Boring 'stuff'?

We are told by the government that health is a key issue for voters, and that the economy is what drives us all, but bloggers show their interest – or lack of it – by the words they *don't* use, as well as the words they do. So it's striking that words like 'health', 'development', 'study', 'investment', 'financial', 'council', 'application' sector are all very low down in a blogger's choice of vocabulary. Perhaps they just can't relate to the language of health, finance and local government?

The rise of bloglish?

So, is there such a thing as 'bloglish', and is it worse or better than other sorts of writing? Some have suggested that bloglish might be an inferior type of English – with more spelling mistakes and careless, unimaginative writing. In fact, our research has found that bloggers have a broadly similar range of vocabulary to other writers, and their level of misspellings and typos is not much higher. This suggests that most bloggers take their role as writers pretty seriously, and they're also talking about serious topics as well: the level of political engagement, and discussion of contemporary politics is evidenced by the prevalence of key proper names such as 'Iraq', 'Bush', and so on.

Miniscule mistakes?

But blogs are not professionally edited and as such they offer excellent evidence of the changing way that people are using the language. For dictionary writers, blogs are a prime resource for tracing the most recent changes in the language; and there is no doubt that some of these trends will find their way into Oxford Dictionaries in the future. For example, we're currently looking at the word 'minuscule': many people use the incorrect spelling 'miniscule', and in Oxford Dictionaries we currently alert people to this being wrong. But looking at the Oxford English Corpus as a whole, the 'i' form is more common and in the blog section it swamps the 'u' spelling – there are more than twice as many 'i' spellings.

The blogs, the corpus, the dictionaries...and the future

The story of the Oxford English Corpus sits at the heart of our ability to track language and reflect real language usage – by real speakers – in our dictionaries. The Oxford English Corpus is a massive text collection of over 1.8 billion words of real 21st-century English. Analysing it and recording the findings in our dictionaries form part of our ongoing commitment as your guide to the English language, and that's why we're highlighting the Corpus on the cover of every one of our English dictionaries:



Did you know? The blog log...

- There are now over 70 million blogs worldwide
- 120,000 blogs created per day worldwide
- c1.5 million postings are made each day
- Japanese is currently the top language for blogs, with 37%.
- English is second with 36% followed by Chinese, Italian, and Spanish.

Source: <http://technorati.com/weblog/2007/04/328.html>

For more information about the Oxford English Corpus or to speak to an editor please contact Juliet Evans on 01865 353911 or email juliet.evans@oup.com. You can also visit our website at www.askoxford.com/oec