

# Cyber- Neologoliferation

In the age of  
the Internet, the Oxford  
English Dictionary is coming  
face to face with the boundlessness of the English language.

**By James Gleick**

**When I got to John Simpson** and his band of lexicographers in Oxford earlier this fall, they were working on the P's. *Pletzel, plish, pod person, point-and-shoot, polyamorous* — these words were all new, one way or another. They had been plowing through the P's for two years but were almost done (except that they'll never be done), and the Q's will be "just a twinkle of an eye," Simpson said. He prizes patience and the long view. A pale, soft-spoken man of middle height and profound intellect, he is chief editor of the Oxford English Dictionary and sees himself as a steward of tradition dating back a century and a half. "Basically it's the same work as they used to do in the 19th century," he said. "When I started in 1976, we were still working very much on these index cards, everything was done on these index cards." He picked up a stack of 6-inch-by-4-inch slips and riffled through them. A thou-

Typography by Sam Winston



sand of these slips were sitting on his desk, and within a stone's throw were millions more, filling metal files and wooden boxes with the ink of two centuries, words, words, words.

But the word slips have gone obsolete now, as Simpson well knows. They are treeware (a word that entered the O.E.D. in September as "computing slang, freq. humorous"). *Blog* was recognized in 2003, *dot-commer* in 2004, *metrosexual* in 2005 and the verb *Google* last June. Simpson has become a frequent and accomplished Googler himself, and his workstation connects to a vast and interlocking set of searchable databases, a better and better approximation of what might be called All Previous Text. The O.E.D. has met the Internet, and however much Simpson loves the O.E.D.'s roots and legacy, he is leading a revolution, willy-nilly — in what it is, what it knows, what it sees.

The English language, spoken by as many as two billion people in every country on earth, has entered a period of ferment, and this place may be the best observation platform available. The perspective here is both intimate and sweeping. In its early days, the O.E.D. found words almost exclusively in books; it was a record of the formal written language. No longer. The language upon which the lexicographers eavesdrop is larger, wilder and more amorphous; it is a great, swirling, expanding cloud of messaging and speech: newspapers, magazines, pamphlets; menus and business memos; Internet news groups and chat-room conversations; and television and radio broadcasts.

The O.E.D. is unlike any other dictionary, in any language. Not simply because it is the biggest and the best, though it is. Not just because it is the supreme authority. (It wears that role reluctantly: it does not presume, or deign, to say that any particular usage or spelling is correct or incorrect; it aims merely to capture the language people use.) No, what makes the O.E.D. unique is a quality for which it can only strive: completeness. It wants every word, all the lingo: idioms and euphemisms, sacred or profane, dead or alive, the King's English or the street's. The O.E.D. is meant to be a perfect record, perfect repository, perfect mirror of the entire language.

James Murray, the editor who assembled the first edition through the final decades of the 19th century, was really speaking of the *language* when he said, in 1900: "The English Dictionary, like the English Constitution, is the creation of no one man, and of no one age; it is a growth that has slowly developed itself adown the ages." And developing faster nowadays. The O.E.D. tries to grasp the whole arc of an ever-changing history. Murray knew that with "adown" he was using a word that could be dated back to Anglo-Saxon of the year 975. When John Updike begins his *New Yorker* review of the new John le Carré novel by saying, "Hugger-mugger is part of life," it is the O.E.D. that gives us the first recorded use of the word, in 1529 ("... not always whyspered in hukermoker," Sir Thomas More) and 27 more quotations from four different centuries. But when *The New York Times* prints a timely editorial about "sock puppets," meaning false identities assumed on the Internet, the O.E.D. has more work to do.

The version now under way is only the third edition. The first, containing 414,825 words in 10 weighty volumes, was presented to King George V and President Coolidge in 1928. Several "supplements" followed, but not till 1989 did the second edition appear: 20 volumes, totaling 21,730 pages. It weighed 138 pounds.

The third edition is a mutation. It is weightless, taking its shape in the

---

*James Gleick, the author, most recently, of "Isaac Newton," is working on a book about the history of information.*

digital realm. To keyboard it, Oxford hired a team of 150 typists in Florida for 18 months. (That was before the verb *keyboard* had even found its way in, as Simpson points out, not to mention the verb *outsource*.) No one can say for sure whether O.E.D.3 will ever be published in paper and ink. By the point of decision, not before 20 years or so, it will have doubled in size yet again. In the meantime, it is materializing before the world's eyes, bit by bit, online. It is a thoroughgoing revision of the entire text. Whereas the second edition just added new words and new usages to the original entries, the current project is researching and revising from scratch — preserving the history but aiming at a more coherent whole.

The revised installments began to appear online in the year 2000. Simpson chose to begin the revisions not with the letter A but with M. Why? It seems the original O.E.D. was not quite a seamless masterpiece. Murray did start at A, logically, and the early letters show signs of the enterprise's immaturity. The entries in A tended to be smaller, with different senses of a word crammed together instead of teased lovingly apart in subentries. "It just took them a long time to sort out their policy and things," Simpson says, "so if we started at A, then we'd be making our job doubly difficult. I think they'd sorted themselves out by. ..." He stops to think. "Well, I was going to say D, but Murray always said that E was the worst letter, because his assistant, Henry Bradley, started E, and Murray always said that he did that rather badly. So then we thought, Maybe it's safe to start with G, H. But you get to G and H, and there's I, J, K, and you know, you think, well, start after that."

So the first wave of revision encompassed 1,000 entries from *M* to *maburat*. The rest of the M's, the N's and the O's have followed in due course. That's why, at the end of 2006, John Simpson and his lexicographers are working on the P's. Their latest quarterly installment, in September, covers *pleb* to *Pomak*. Simpson mentions rather proudly that they scrambled at the last instant to update the entry for *Pluto* when the International Astronomical Union voted to rescind its planethood. *Pluto* had entered the second edition as "1. A small planet of the solar system ... " discovered in 1930 and "2. The name of a cartoon dog ..." first appearing in 1931. The Disney meaning was more stable, it turns out. In O.E.D.3, Pluto is still a dog but merely "a small planetary body."

**E**ven as they revise the existing dictionary in sequence, the O.E.D. lexicographers are adding new words wherever they find them, at an accelerating pace. Beside the P's, September's freshman class included *agroterrorism*, *bahookie* (a body part), *beer pong* (a drinking game), *bippy* (as in, you bet your —), *chucklesome*, *cypherpunk*, *tuneage* and *wonkey*. Every one of these underwent intense scrutiny. The addition of a new word is a solemn matter.

"Because it's the O.E.D.," says Fiona McPherson, a new-words editor, "once something goes in, it cannot ever come out again." In this respect,

you could say that the O.E.D. is a *roach motel* (added March 2005: "Something from which it may be difficult or impossible to be extricated"). A word can go *obs.* or *rare*, but the editors feel that even the most ancient and forgotten words have a way of coming back — people rediscover them or reinvent them — and anyway, they are part of the language's history.

The new-words department, where that history rolls forward, is not to everyone's taste. "I love it, I really really love it," McPherson says. "You're at the cutting edge, you're dealing with stuff that's not there and you're, I suppose, shaping the language. A lot of people are more interested in the older stuff; they like nothing better than reading through 18th-century texts looking for the right word. That doesn't suit me as much, I have to say." *Cutting edge*, incidentally, is not a new word: according to the O.E.D., H. G. Wells used it in its modern sense in 1916.

As a rule, a neologism needs five years of solid evidence for admission to the canon. "We need to be sure that a word has established a reasonable amount of longevity," McPherson says. "Some things do stick around that you would never expect to stick around, and then other things, you think *that* will definitely be around, and everybody talks about it for six months, and then. ..."

Still, a new word as of September is *bada-bing*: American slang "suggesting something happening suddenly, emphatically, or easily and predictably." "The Sopranos" gets no credit. The historical citations begin with a 1965 audio recording of a comedy routine by Pat Cooper and continue with newspaper clippings, a television news transcript and a line of dialogue from the first "Godfather" movie: "You've gotta get up close like this and bada-bing! you blow their brains all over your nice Ivy League suit." The lexicographers also provide an etymology, a characteristically exquisite piece of guesswork: "Origin uncertain. Perh. imitative of the sound of a drum roll and cymbal clash.... Perh. cf. Italian *bada bene* mark well."

But is *bada-bing* really an official part of the English language? What makes it a word? I can't help wondering, when it comes down to it, isn't *bada-bing* (also *badda-bing*, *badda badda bing*, *badabing*, *badaboom*) just a noise?

"I dare say the thought occurs to editors from time to time," Simpson says. "But from a lexicographical point of view, we're interested in the conventionalized representation of strings that carry meaning. Why, for example, do we say *Wow!* rather than some other string of letters? Or *Zap!* Researching these takes us into interesting areas of comic-magazine and radio-TV-film history and other related historical fields. And it often turns out that they became institutionalized far earlier than people nowadays may think."

When Murray began work on O.E.D.1, no one had any idea how many words were there to be found. Probably the best and most comprehensive dictionary of English was American, Noah Webster's: 70,000 words. That number was a base line. Where were the words to be discovered? For the first editors it went almost without saying that the source, the wellspring, should be the *literature* of the language. Thus it began as a dictionary of

the written language, not the spoken language. The dictionary's first readers combed Milton and Shakespeare (still the single most quoted author, with more than 30,000 references), Fielding and Swift, histories and sermons, philosophers and poets. "A thousand readers are wanted," Murray announced in his famous 1879 public appeal. "The later 16th-century literature is very fairly done; yet here several books remain to be read. The 17th century, with so many more writers, naturally shows still more unexplored territory." He considered the territory to be large, but ultimately finite.

It no longer seems finite.

"We're painting the Forth Bridge!" says Bernadette Paton, an associate editor. "We're running the wrong way on a travolator!" (I get the first part — "allusion to the huge task of maintaining the painted surfaces of the railway bridge over the Firth of Forth" — but I have to ask about *travolator*. Apparently it's a moving sidewalk.)

The O.E.D. is a historical dictionary, providing citations meant to show the evolution of every word, beginning with the earliest known usage. So a key task, and a popular sport for thousands of volunteer word aficionados, is antedating: finding earlier citations than those already known. This used to be painstakingly slow and chancy. When Paton started in new words, she found herself struggling with *headcase*. She had current citations, but she says she felt sure it must be older, and books were of little use. She wandered around the office muttering *headcase*, *headcase*, *headcase*. Suddenly one of her colleagues started singing: "My name is Bill, and I'm a *headcase*! They practice making up on my face." She perked up.

"What date would that be?" she asked.

"I don't know, it's a Who song," he said, "1966 probably, something like that."

So "I'm a Boy," by P. Townshend, became the O.E.D.'s earliest citation for *headcase*.

Antedating is entirely different now: online databases have opened the floodgates. Lately Paton has been looking at words starting with *pseudo-*. Searching through databases of old newspapers and historical documents has changed her view of them. "I tended to think of *pseudo-* as a prefix that just took off in the 60's and 70's, but now we find that a lot of them go back much earlier than we thought." Also in the P's, *poison pen* has just been antedated with a 1911 headline in The Evening Post in Frederick, Md. "You get the sense that this sort of language seeps into local newspapers first," she says. "We would never in a million years have sent a reader to read a small newspaper like that."

The job of a new-words editor felt very different precyberspace, Paton says: "New words weren't proliferating at quite the rate they have done in the last 10 years. Not just the Internet, but text messaging and so on has created lots and lots of new vocabulary." Much of the new vocabulary appears online long before it will make it into books. Take *geek*. It was not till 2003 that O.E.D.3 caught up with the main modern sense: "a person who is extremely devoted to and knowledgeable about computers or related technology." Internet chitchat provides the earliest known reference, a posting to a Usenet newsgroup, net.jokes, on Feb. 20, 1984.

The scouring of the Internet for evidence — the use of cyberspace as a language lab — is being systematized in a program called the Oxford English Corpus. This is a giant body of text that begins in 2000 and now contains more than 1.5 billion words, from published material but also from Web sites, Weblogs, chat rooms, fanzines, corporate home pages

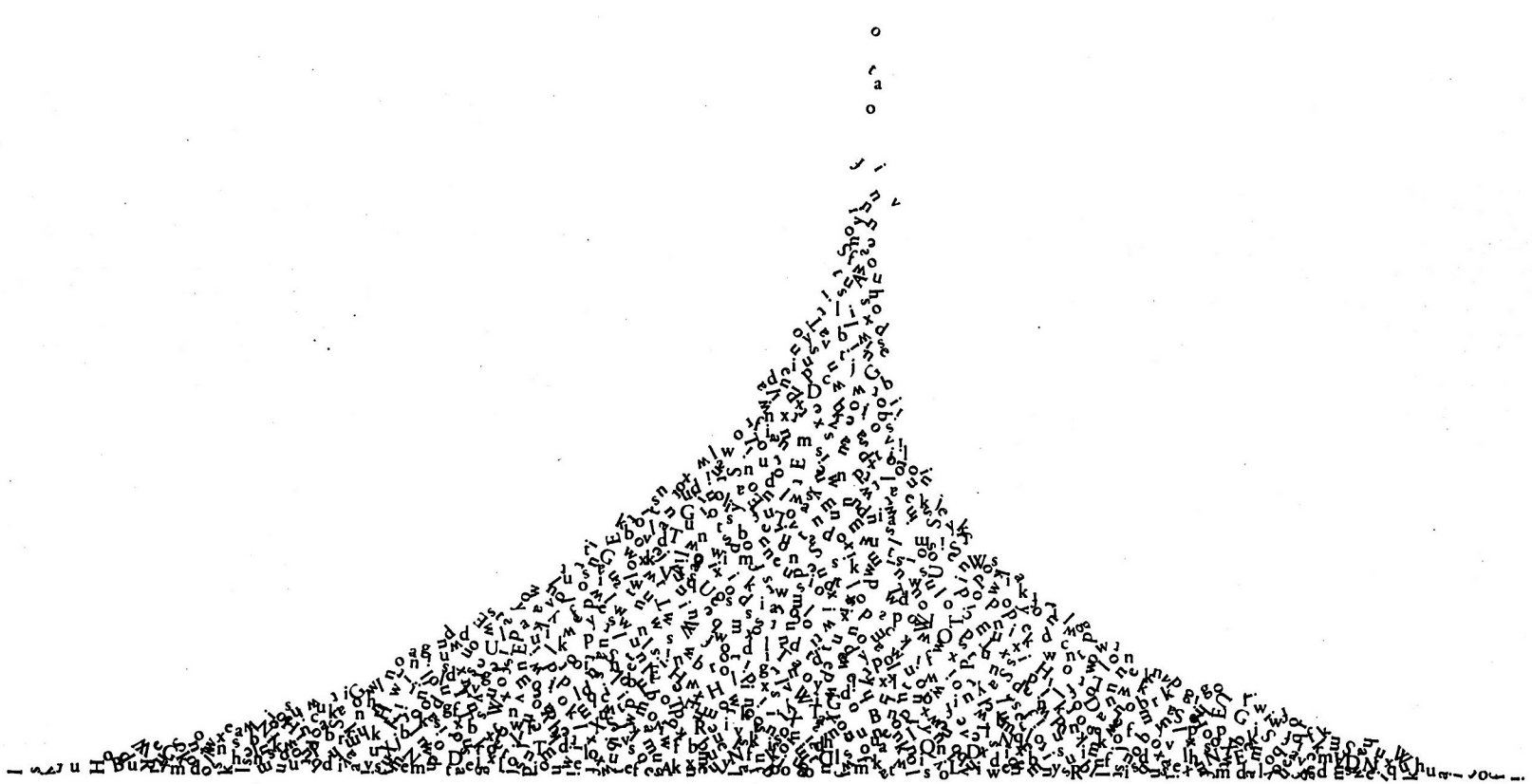
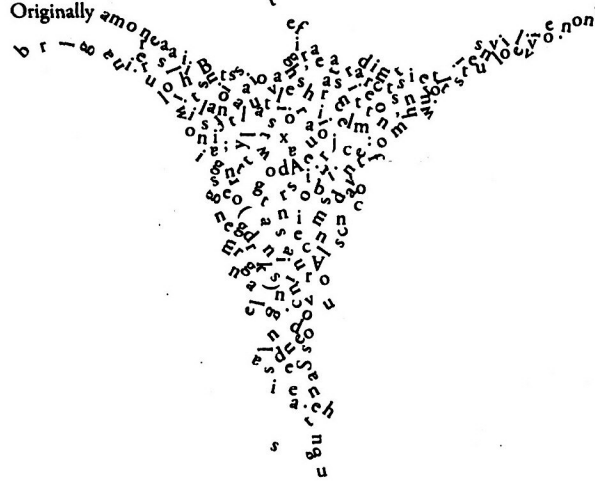
To keyboard the new O.E.D., Oxford hired a team of 150 typists in Florida for 18 months. (That was before the verb *keyboard* had even found its way in, not to mention the verb *outsource*.)

**Zampone** In Italian cookery: a pig's trotter, cured with spiced ground pork, then cured and dried

**Zhoosh** Originally among the Chechens, a soup made from the meat of any one species

**Zip**

**Zombie**



and radio transcripts. The corpus sends its home-built Web crawler out in search of text, raw material to show how the language is really used.

I'm too embarrassed to ask the lexicographers if they have a favorite word. They get that a lot. Peter Gilliver tells me his anyway: *twiffler*. A twiffler, in case you didn't know, is a plate intermediate in size between a dinner plate and a bread plate. "I love it because it fills a gap," Gilliver says. "I also love it because of its etymology. It comes from Dutch, like a lot of ceramics vocabulary. *Twijfelaar* means something intermediate in size, and it comes from *twijfelen*, which means to be unsure. It's a plate that can't make up its mind!"

Fiona McPherson gives me *mondegreen*. A mondegreen is a misheard lyric, as in, "Lead on, O kinky turtle." It is named after Lady Mondegreen. There was no Lady Mondegreen. The lines of a ballad, "They hae slain the Earl of Murray, / And laid him on the green" are misheard as "They hae slain the Earl of Murray and Lady Mondegreen."

"A lot of people are just really excited by that word because they think it's amazing that there is a word for that concept," McPherson says.

I have my own favorites among the newest entries in O.E.D.3. *Pixie dust* is, as any child knows, "an imaginary magical substance used by pixies." *Air kiss* is defined with careful anatomical instructions plus a note: "sometimes with the connotation that such a gesture implies insincerity or affectation." *Builder's bum* is reportedly Brit. and colloq., "with allusion to the perceived propensity of builders to expose inadvertently this part of the body."

It is clear that the English of the O.E.D. is no longer the purely written language, much less a formal or respectable English, the diction recommended by any authority. Gilliver, a longtime editor who also seems to be the O.E.D.'s resident historian, points out that the dictionary feels obliged to include words that many would regard simply as misspellings. No one is particularly proud of the new entry as of December 2003 for *nucular*, a word not associated with high standards of diction. "Bizarrely, I was amazed to find that the spelling n-u-c-u-l-a-r has decades of history," Gilliver says. "And that is not to be confused with the quite different word, *nucular*, meaning 'of or relating to a nucle.'"  
There is even a new entry for *miniscule*; it has citations going back more than 100 years.

Yet the very notion of correct and incorrect spelling seems under attack. In Shakespeare's day, there was no such thing: no right and wrong in spelling, no dictionaries to consult. The word *debt* could be spelled det, dete, dett, dette or dept, and no one would complain.

Then spelling crystallized, with the spread of printing. Now, with mass communication taking another leap forward, spelling may be diversifying again, spellcheckers notwithstanding. The O.E.D. so far does not recognize *straight-laced*, but the Oxford English Corpus finds it outnumbering *strait-laced*. Similarly for *just desserts*.

To explain why cyberspace is a challenge for the O.E.D. as well as a godsend, Gilliver uses the phrase "sensitive ears."

"You know we are listening to the language," he says. "When you are listening to the language by collecting pieces of paper, that's fine, but now it's as if we can hear everything said anywhere. Members of some tiny English-speaking community anywhere in the world just happen to commit their communications to the Web: there it is. You thought some word was obsolete? Actually, no, it still survives in a very small community of people who happen to use the Web — we can hear about it."

In part, it's just a problem of too much information: a small number of

lexicographers with limited time. But it's also that the O.E.D. is coming face to face with the language's boundlessness.

The universe of human discourse always has backwaters. The language spoken in one valley was a little different from the language of the next valley and so on. There are more valleys now than ever, but they are not so isolated. They find one another in chat rooms and on blogs. When they coin a word, anyone may hear.

Neologisms can be formed by committee: *transistor*, Bell Laboratories, 1948. Or by wags: *booboisie*, H.L. Mencken, 1922. But most arise through spontaneous generation, organisms appearing in a petrie dish, like *blog* (c. 1999). If there is an ultimate limit to the sensitivity of lexicographers' ears, no one has yet found it. The rate of change in the language itself — particularly the process of neologism — has surely shifted into a higher gear now, but away from dictionaries, scholars of language have no clear way to measure the process. When they need quantification, they look to the dictionaries.

"An awful lot of neologisms are spur-of-the-moment creations, whether it's literary effect or it's conversational effect," says Naomi S. Baron, a linguist at American University, who studies these issues. "I could probably count on the fingers of a hand and a half the serious linguists who know anything about the Internet. That hand and a half of us are fascinated to watch how the Internet makes it possible not just for new words to be coined but for neologisms to spread like wildfire."

It's partly a matter of sheer intensity. Cyberspace is an engine driving change in the language. "I think of it as a saucepan under which the temperature has been turned up," Gilliver says. "Any word, because of the interconnectedness of the English-speaking world, can spring from the backwater. And they are still backwaters, but they have this instant connection to ordinary, everyday discourse." Like the printing press, the telegraph and the telephone before it, the Internet is transforming the language simply by transmitting information differently. And what makes cyberspace different from all previous information technologies is its intermixing of scales from the largest to the smallest without prejudice, broadcasting to the millions, narrowcasting to groups, instant messaging one to one. So anyone can be an O.E.D. author now. And, by the way, many try. "What people love to do is send us words they've invented," Bernadette Paton says, guiding me through a windowless room used for storage of old word slips. *Will you put the word I have invented into one of your dictionaries?* is a question in the AskOxford.com FAQ. All the submissions go into the files, and until there is evidence for some general usage, that's where the wannabes remain. Don't bother sending in FAQ. Don't bother sending in *wannabes*. They're not even particularly new. For that matter, don't bother sending in anything you find via Google. "Please note," the O.E.D.'s Web site warns solemnly, "it is generally safe to assume that examples found by searching the Web, using search engines such as Google, will have already been considered by O.E.D. editors." ■

When you are listening to the language by collecting pieces of paper, that's fine, but now it's as if we can hear everything said anywhere, says one editor.