Language and the Internet
Second Edition
DAVID CRISTAL

CAMBRIDGE UNIVERSITY PRESS
Contents

A note on the second edition vi
Preface to the first edition ix

1 A linguistic perspective 1
2 The medium of Netspeak 7
3 Finding an identity 66
4 The language of e-mail 99
5 The language of chatgroups 118
6 The language of virtual worlds 135
7 The language of the Web 2
8 New varieties 238
9 The linguistic future of the Internet 265

References 277
Index of authors 286
Index of topics 289
New varieties

The first edition of *Language and the Internet* appeared in 2001, with the preface predicting rapid advances in the field, and thus the need for revision on a shorter time-scale than is usual in book publication. The prediction turned out to be accurate, but in unexpected ways. I had anticipated that the four domains of e-mail, chat, virtual worlds, and the Web would evolve in interesting directions; but I did not expect to have to add an additional domain to this list — and certainly not two. Yet that is what has happened. As will already be apparent from chapters 1 and 2, the activities generally referred to as *blogging* and *instant messaging* have developed a linguistic character sufficiently distinctive as to require separate treatment. And other activities, especially in Internet telephony, are becoming apparent just over the linguistic horizon.

Blogging

*Blogs* were already in existence when I was preparing the first edition of this book in 2000, but they had achieved little public presence, and I was not aware of them. Five years on, and the name is everywhere, along with a whole family of derived nomenclature. Those who blog, *bloggers*, carry out the activity of *blogging*, setting up a *blogsite* with a unique web address in order to do so. They may also locate their page within a bigger site, which takes *blog feeds* from many sources, and includes other (*non-blog*) material. As bloggers gain experience, they will compile a *blogroll* listing their favourite links to other blogs, often shown as a sidebar on their screen. If they dart about from one site to another, they are engaging in *blog hopping*. If their blog contains a request to readers (e.g. for contributions), it may be called a *blog* (a 'begging blog'). If a blog goes on for too long, the writer may be described as having *blogorrhea*. If a topic or site attracts a considerable amount of online attention, the result is a *blogstorm* or *blog swarm*. Those who are wise in the ways of blogs are sometimes called the *blogosite* or *blogerati*. The totality of all blogsites is known as the *blogosphere*.

The use of broadband technology has added further dimensions to blogging, and multimedia blogs are now commonplace. A blog which consists chiefly of posts sent by mobile phone (p. 262) or some other wireless device is a *moblog*. If it incorporates a voice recording sent by mobile phone, or a music download (such as an MP3 blog), it would be an *audioblog*. The inclusion of digital photographs would make it a *photoblog*. Adding video clips would turn it into a *videoblog*, or *vlog*. 1 The audio and video dimensions are bound to foster novel linguistic developments, in due course, but it is too soon (in mid-2005) to say what character these might take.

*Blog* was declared 'word of the year' by the dictionary publishers Merriam-Webster in 2004. It is an abbreviation of *weblog* or *weblog*, a term which was first used in 1997, when Jorn Barger coined it to refer to his Robot Wisdom Weblog. 2 In mid-1999, the shortened form appeared, when Peter Merholz used the phrase *we blog* in a sidebar on his site, and the usage quickly evolved as an independent noun and verb. 3 Later that year the first sites to offer hosted weblog tools arrived, notably *Blogger*. 4 As the full form of

---

1 The first *VloggerCon* was held in January 2005: http://vloggercon.blogspot.com/
3 See http://www.peterme.com/archives/00000205.html; http://www.bradlands.com/weblog/1999-09.shtml; September#92010;#9201999. The verb usage also includes a simultaneous present-tense usage, as in *I am blogging the match* — that is, the blogger is posting reports or comments about the match as it takes place.
4 Launched by Pyra Labs, it attracted over a million users by 2003. It was bought by Google in 2004. Other blog-hosting services include *GreatestJournal* (www.greatestjournal.com), *LiveJournal* (www.livejournal.com), *Xanga* (www.xanga.com), and *Pitas* (www.pitas.com). More advanced tools are available on some servers for those wanting to develop more sophisticated sites, such as for corporate or school use.
the name suggests, a blog is a web application which allows the user to enter, display, and edit posts at any time. It is essentially a content-management system—a way of getting content onto a web page. Most users think of blogging as a genre akin to diary-writing or bulletin-posting, and add posts with some regularity, usually daily, often several times a day. The posts then appear on the site in chronological order, identified by date and time, typically with the most recent at the top.

Blogs rapidly came to be used for a remarkable range of purposes. At one extreme there is the personal diary, kept by an individual who wants to inform the world of his or her activities, interests, and opinions. At the other extreme, there is the corporate blog, maintained by an institution—such as a radio station, a music store, a university department, a search engine, or a political party—to inform a potential readership of its activities. Larger blogs tend to be multi-authored. Many of them are interactive, welcoming spontaneous or invited feedback about the topic of the blog. Such blogs attract regular readers, and blogging communities have thus emerged, with the resulting exchanges closely resembling those encountered in asynchronous chatgroups (chapter 5). But blogs are different. Most are under the control of a single user, who therefore manages the direction and content of the site (whereas any chatgroup member may start a new theme or thread), and gives it a distinctive personality. They typically lack the intrinsically spontaneous interactivity of chatgroups. And they rely on links to other sites in a way that chatgroups typically do not.

Links are very important. Some blogs consist of little more than a long list of hyperlinks (p. 209) to sites that relate to a particular theme, often with a summary of their content and an evaluation of some kind. Other blogs consist of current and archived messages, with each message given its own static address, or permalink. What makes blogging different from other Internet media is the way the software combines the editorial flexibility required by personal web pages with the dynamic interactivity required when there are so many points of potential connection. The analogy of a blog with a diary is only partly accurate. A diary is a linear sequence of entries following a certain chronology, but a blog is not simply a web page of indefinite length, down which one scrolls to read messages earlier in time. Anyone who restricted a blog in this way would be seriously underusing its potential functionality. The various kinds of linkage place a blog within a network of related content that the blogger (and the blog’s readers) are invited to explore. The more a site is linked to others, the more presence it has, and thus the more influence and authority. Special tools help users to explore efficiently, such as servers which let it be known when a blog has been updated.5

The personal element in blogging, and its lack of any centralizing or moderating control, soon made it attractive to those holding radical or extreme opinions, who would otherwise have found it difficult to get their views into the public domain. But the way the medium facilitated online presence motivated people of all persuasions to contribute to issues of the moment. Among the earliest blogstorms were debates in the USA surrounding the war on terrorism, especially after the events of 11 September 2001, followed by an even larger storm relating to the invasion of Iraq, both for and against (‘war bloggers’). People were quick to see the potential of blogging as a new form of public awareness, combining elements of journalism, advertising, public relations, outreach, consumer news, propaganda, and interest-group formation, as well as providing an easy means of tapping into public opinion and obtaining feedback. The medium was thus soon populated by all kinds of professional pundits and commentators, including some leading politicians and journalists, with some blogsites achieving ‘must read’ status. At the same time, it proved just as attractive to those whose voices were not usually heard, such as Iraqi civilians and serving

5 An example is Blogrolling (http://www.blogrolling.com/). The automatic notification to a site that there has been an update or response to a blog posting is known as trackback. See the account at http://www.movabletype.org(trackback/beginners.)
New varieties contrast, there are many thematic blogs restricted to a single subject domain or topic of interest, such as those which focus on a theme within religion, law (‘blawgs’), politics, software, news, sport, or hobbies. Some blogs are simply bibliographic in character, providing a list of useful sites on a particular topic. Most are collaborative, allowing contributions from all who wish to participate, and these give the medium the character of an asynchronous chatgroup (chapter 6). Indeed, some argue that blogging will become increasingly interactive, as its technology becomes more widespread, and will eventually replace a great deal of what goes on in chatgroup situations. This is already clearly the case in the collaborative blogs associated with traditional media, such as a newspaper or a radio station. Listeners to a radio programme can be sent to an associated blog in order to ‘continue the conversation’ after the programme is over. Collaboration can also be seen in consumer watchdog blogs, which monitor products and provide evaluations, and in advice blogs, which deal with problems in the manner of a newspaper ‘agony aunt’, but with the difference that advice may come in from all directions.

It is too early to say what the social consequences of blogging are: the debate is lively and ongoing. Plainly there are fresh implications for any discussion of the role of the media in society, the nature of political debate and empowerment, the quality of news coverage, and legal questions to do with intellectual property and freedom of speech. Similarly, given the speed at which the technology is developing, it is too early to say anything definite about the typical visual organization of a blog or its functionality. But some of the linguistic consequences of this new medium are already clear, as these extracts illustrate (both written in June 2005). Each is a whole entry, with its own date and time.

I just watched a most wonderful TV film this evening: some not too original story about a man and a woman who are Mr and Mrs Right, but don’t know, go through a lot of confusion and tangle and complexity and unhappiness — to, of course, finally get each other. Not too original, really, not at all, but quite wittily
and heart-warmingly done – or perhaps I was just in a resonating and receptive mood. Anyway – I liked it and it made me feel good afterwards (which is my main criterium on whether a film is worth watching or not).

Went out, stopped by the Mall to get some coffee – problem as usual with all the car parking, and as if that wasn’t enough the lights controlling the way into the car-park, you know the twostage green and red switching thingies weren’t working so it took ages. And there’s never mechanics around when you want them, is there – tho I did see someone in a uniform lurking about – but come to think of it what would you need a uniform for to fix lights, so maybe they were just there for something else.

Anyways, it took me an hour instead of what I was expecting, 10 minutes.

Here we have examples of a style of writing which has never been seen in a public, printed form, outside of literature, and even there it would take an ingenious novelist indeed to capture its innocent spontaneity and unpredictable thematic direction. It is difficult to know how to describe the style, because it falls uneasily between standard and non-standard English. Both extracts illustrate writing which is largely orthodox with respect to the main dimensions that identify standardness – spelling, punctuation, and grammar, but they depart from the norms in various ways. In spelling, we see controlling for controlling, tho for though, and two-stage for two-stage; in punctuation, we see the avoidance (in the second extract) of apostrophes. In vocabulary, we see criterium for criterion and anyways for anyway. There are several features of informal written English which would be eliminated in a copy-edited version of such texts for publication, such as the unconstrained use of the dash to mark a change in the thought or the use of commas to follow the rhythm of the ‘speaker’. As regards grammar, stretches of text defy conventional grammatical analysis in terms of sentences. The discourse expresses a sequence of units of thought, but these do not correspond to the kinds of

sentence division we have been taught to associate with ‘elegant’ writing. At lower levels of grammar, too, there are features which would be considered unacceptable in traditionally printed publications, such as to, of course, finally get (which atomically splits the infinitive), there’s never mechanics (breaking a concord rule), and criterium on (wrong preposition). In its unconstrained flow, it is – I imagine – as close to the way the writers talk as it is possible to get. Certainly, the style drives a coach and horses through everything we would be told in the grammatical tradition of the past 250 years about how we should write.

Before the emergence of standard English, of course, such a style of writing would not have attracted any notice at all. If we examine the spontaneous letter-writing of the late Middle Ages, for example, or some of the manuscript accounts of law-court proceedings, we will find a very similar set of features (see fn. 8 above). It is a style which was once the norm, for all kinds of writing, but which gradually went out of public use once the standard language was institutionalized in manuals of grammar, punctuation, and usage, beginning in the second half of the eighteenth century. It was finally eliminated when publishers developed copy-editing procedures to ensure that their newspapers, magazines, and books conformed to an in-house style. After that point, it was virtually impossible to see anything in print which had not been through a standardizing process. The text that you are reading now has been altered, in many tiny respects, from the text that I originally wrote. If I send a letter to a newspaper, or write a review for a magazine, my language ceases to be under my control, and is subject to an editorial process which can change its character noticeably. And this is why blogging is so significant. Only here do we have the opportunity to see written discourses of sometimes substantial length which have had no such editorial interference. It is written language in its most naked form.

---

Footnote:

8 For the notion of standard English and its relation to non-standard, along with a number of illustrations of texts, see Crystal (2004).

9 It would be premature to hazard a generalization about blog length. People who update their blogs at the end of a day, or each week, are likely to write much longer entries than those who do so several times a day. The above entries are around 100
I am not of course suggesting that these extracts are typical of the entire genre. All kinds of people blog, and all kinds of style are present, from informal to formal, from non-standard to standard. Some bloggers are scrupulous about formality and consistency, and would be mortified if their text appeared in the blogosphere with a missing apostrophe. And the increase in corporate blogs means that there will be an increasing element of editorial control, leading in due course to greater standardization. Some blogs are no more than links to sites, such as newspaper articles, which have been edited in the traditional way. But personal blogs do illustrate something that is not found elsewhere today: a variety of writing intended for public consumption which appears exactly as the author wrote it, which is not constrained by other genre conventions, and which privileges linguistic idiosyncrasy. I call it, on analogy with free verse, free prose.

The statistics of blogging are notoriously uncertain. Estimates of the number of active blogs in 2003 varied from 1 million to 3 million. Commentators were suggesting that the number of blogs was doubling every six months. Estimates in 2005 were typically in the region of 12 to 15 million, but some were as high as 50 million worldwide. In April 2005, Microsoft reported that over 4.5 million ‘spaces’ (i.e. blogs) had been created since January of that year. Updates were running at around 170,000 a day. Of course, by no means all of these new sites remain active: only about one new blogger in five becomes a regular user of the medium. And of the sites which are active, it is thought that only about 20 per cent attract traffic. It seems, sadly, that nobody reads the vast majority of items that are ‘out there’ – except possibly the occasional blogosphere-travelling linguist. On the other hand, a popular blog attracts a huge readership, often over a million a month, which is more than the audience reached by many newspapers. The French music station, Skyblog, was reporting 1.6 million users in 2005 – 2.6 per cent of the population of France.

Blogging has introduced a new era of interactivity to websites. It has already provided a new means of publishing and distributing information on the Internet, and as the technology develops it could be the next stage in the evolution of online chat. There will be implications for the character of e-mail, too, if bloggers increasingly use the option to have their updates sent out as mail. And the multimedia developments in audio and video blogging will inevitably have consequences, in the form of new linguistic conventions, as usage grows.

**Instant messaging**

The exchange of messages online is a repeated theme of the earlier chapters of this book; but each method has its limitations. Chatgroups exist in real time, but have an open membership of typically anonymous participants, so are not appropriate for intimate exchanges between couples or small groups who know each other well. A blog can be used for a small circle of contacts, but it is not in real time. Even with e-mails or text-messaging, there is an inevitable lag between the sending and receiving of messages (p. 34), and there is no way of knowing whether the person you are sending the message to is there or available to receive it. In any case, an e-mail exchange is not really a conversation, in the sense of an exchange that can be carried on over an indefinite period of time. Each e-mail message involves a series of steps to send and receive, which would become absurd if the medium were used for the kind of feedback exchanges illustrated below. For anyone who wanted to use a computer to carry on a natural conversation online in real time with another person, or a small group of friends or acquaintances, there was nothing that would allow them to do it – until instant messaging

---

10 For the statistics above, see [http://dijest.com/bci/](http://dijest.com/bci/). Blogstreet monitors over 100,000 blogs, and publishes lists, such as the top 100 most popular blogs: [http://www.blogstreet.com/top100.html](http://www.blogstreet.com/top100.html). The MIT Media Laboratory uses Blogdex to track the way information spreads through the blog community: see [http://blogdex.net](http://blogdex.net).