A linguistic revolution

by David Crystal

A linguist can't help but be impressed by the Internet. It is an extraordinarily diverse medium, holding a mirror up to many sides of our linguistic nature. The World Wide Web, in particular, offers a home to virtually all the styles which have so far developed in the written language – newspapers, scientific reports, bulletins, novels, poems, prayers – you name it, you'll find a page on it. Indeed, it is introducing us to styles of written expression which none of us have ever seen before. It has often been said, the Internet is a social revolution – yes, indeed, but it seems to me that it is also a linguistic revolution.

I recognize that the Internet is not a single thing, but consists of several domains - e-mails, the World Wide Web, chatrooms (those which exist in real time and those which do not), the world of fantasy games, instant messaging, blogging, and more. Each offers us novel possibilities of human communication which I think can genuinely be called revolutionary. And everything is moving very fast. In 2001 the first edition of my *Language and the Internet* (CUP) didn't even mention instant messaging and blogging. In the second edition, out in 2006, they form a separate chapter. So much can happen in five years!

In e-mails, what is revolutionary is not the way some of its users are cavalier about their typing accuracy, permitting misspellings, and omitting capitalization and punctuation. This is a rather minor effect, which rarely interferes with intelligibility. It is patently a special style arising out of the pressures operating on users of the medium, plus a natural desire (especially among younger - or younger-minded - users) to be idiosyncratic and daring. And that is how it is perceived. If I receive an e-mail from Smith in which he mis-spells a word, I do not conclude from this that 'Smith can't spell'. I simply conclude that he was in a hurry. I know this because I do the same thing myself, when I am in a hurry. There is nothing truly revolutionary here.

What is revolutionary about e-mails is the way the medium permits what is called *framing*. You receive a message which contains, say, three different points in a single paragraph. You can, if you want, reply to each of these points by taking the paragraph, splitting it up into three parts, and then responding to each part separately, so that the message you send back then looks a bit like a play dialogue. Then, your sender can do the same thing to your responses, and when you get the message back, you see his replies to your replies. You can then send the lot on to someone else for further comments, and when it comes back, there are now three voices framed on the screen. And so it can go on – replies within replies within replies - and all unified within the same screen typography. There's never been anything like this in the history of human written communication.

The pages of the Web offer a different kind of revolutionary development. The one thing we can say about traditional writing is that it is permanent. You open a book at page 6, close the book, then open it at page 6 again. You expect to see the same thing. You would be more than a little surprised if the page had changed in the interim. But this kind of impermanence is perfectly normal on the Web - where indeed you can see the page changing in front of your eyes. Words appear and disappear, in varying colours. Sentences slide onto the screen and off again. Letters dance around. The Web is truly part of a new, animated linguistic channel – more dynamic than traditional writing, and more permanent than traditional speech. It is neither speech nor writing. It is part of a new medium.

Real-time Internet discussion groups - chatrooms - also offer a revolutionary set of possibilities. You see on your screen messages coming in from all over the world. If there are 30 people in the room, then you could be seeing 30 different messages, all making various contributions to the theme, but often clustering into half a dozen or more sub-conversations. It's like being in a cocktail party where there are other conversations going on all around you. In the party, of course, you can't pay attention to them. In a chatroom you can't avoid them. It has never been possible before, in the history of human communication, to 'listen' to 30 people at once. Now you can. Moreover, you can respond to as many of them as your mental powers and typing speed permit. This too is a revolutionary state of affairs, as far as speech is concerned.

But there's a further reason for the revolutionary status of the Internet - the fact that it offers a home to *all* languages – as soon as their communities have a functioning computer technology, of course. Its increasingly multilingual character has been the most notable change since it started out – not very long ago - as a totally English medium. There's a story the former US vice-president Al Gore tells. He was reporting the remark of the 8-year-old son of Kyrgyzstan's President Akayev, who told his father that he had to learn

English. When asked why, the child apparently replied: 'Because, daddy, the computer speaks English.'

For many, indeed, the language of the Internet 'is' English. There was a headline in *The New York Times* in 1996 which said simply: 'World, Wide, Web: 3 English Words'. The article, by Michael Specter, went on to say: 'if you want to take full advantage of the Internet there is only one real way to do it: learn English'. He did acknowledge the arrival of other languages: 'As the Web grows', he said, 'the number of people on it who speak French, say, or Russian will become more varied and that variety will be expressed on the Web. That is why it is a fundamentally democratic technology', he said, 'but it won't necessarily happen soon.'

Well, the evidence is growing that this conclusion was wrong. With the Internet's globalization, the presence of other languages has steadily risen. By the mid-1990s, a widely quoted figure was that about 80% of the Net was in English. This figure derived from the first major study of language distribution on the Internet, carried out in 1997 by Babel, a joint initiative of the Internet Society and Alis Technologies. This showed English well ahead, but with several other languages entering the ring – notably German, Japanese, French, and Spanish.

Since then, the estimates for English have been steadily falling. Some commentators are now predicting that before long the Web (and the Internet as a whole) will be predominantly *non*-English, as communications infrastructure develops in Europe, Asia, Africa, and South America. A recent Global Reach survey estimated that people with Internet access in non-English-speaking countries increased between 1995 and 2000 from 7 million to 136 million. In 1998, there was another surprise: the number of newly created Web sites *not* in English passed the total for newly created sites that *were* in English. And at a conference on Search Engine Strategies in London in 2000, a representative of Alta Vista was predicting that by the end of 2002 less than 50% of the Web would be in English. This has turned out to be true. In certain parts of the world, the local language is already dominant. According to one Japanese Internet author, Yoshi Mikami, 90% of Web pages in Japan are already in Japanese.

Spend an hour hunting for languages on the World Wide Web and you'll find hundreds. In 2001 I spent a few days tracking down as many examples as I could find, for my book *Language and the Internet*. I found one site, called World Language Resources, which lists products for 728 languages. I found an African resource list which covered several local languages; Yoruba, for example, was illustrated by some 5000 words, along with proverbs, naming patterns, and greetings. Another site dealt with no less than 87 European minority

languages. Some of the sites were very small in content, of course, but nonetheless extensive in range: one gave the Lord's Prayer in nearly 500 languages.

Nobody has yet worked out just how many languages have obtained a modicum of presence on the Web. I found over 1000 quite quickly. It's not difficult to find evidence of a Net presence for all the more frequently used languages in the world, and for a large number of minority languages too. I'd guess that about a quarter of the world's languages – that's about 1500 - have some sort of cyber existence now.

It's important to point out that in all these examples I'm talking about language presence in a real sense. These aren't sites which only analyse or talk about languages, from the point of view of linguistics or some other academic subject. They're sites which allow us to see languages as they are. In many cases, the total Web presence, in terms of number of pages, is quite small. The crucial point is that the languages are out there, even if they're represented by only a sprinkling of sites.

The Internet is the ideal medium for minority languages. If you are a speaker or supporter of an endangered language – an aboriginal language, say, or one of the Celtic languages – you're keen to give the language some publicity, to draw its plight to the attention of the world. Previously, this was very difficult to do. It was hard to attract a newspaper article on the subject, and the cost of a newspaper advertisement was prohibitive. It was virtually impossible to get a radio or television programme devoted to it. But now, with Web pages and e-mail waiting to be used, you can get your message out in next to no time, in your own language – with a translation as well, if you want - and in front of a global audience whose potential size makes traditional media audiences look minuscule by comparison. Chat rooms, moreover, are a boon to speakers living in isolation from each other, as now there can be a virtual speech community to which they can belong.

On the other hand, I have to recognize that developing a significant cyber-presence for a language is not easy. There's a sort of 'critical mass' of Internet penetration which has to build up in a country, before a language develops a vibrant cyber-life. It's not much use, really, to have just one or two sites in a local language on the Web. People wanting to use or find out about the language would soon get bored. The number of sites has to build up until, suddenly, everybody's using them and adding to them and talking about them. That's a magic moment, and only a few hundred languages have so far reached it. In the jargon of the Internet, there needs to be lots of good 'content' in the local languages out there, and until there is, people will stay using the languages that have managed to accumulate content – English, in particular.

So the character of a multilingual Internet isn't entirely clear. It will all depend on how quickly new sites can build up a local language momentum. There are also a number of practical difficulties. Until quite recently there were real problems in using the characters of the keyboard to cope with the alphabetical diversity of the world's languages. Because it was the English alphabet that was the standard, only a very few non-English accents and diacritics could be handled. If it was a foreign word with some strange-looking accent marks, the Internet software would simply ignore them, and assume they weren't important. This can still happen – but things have moved on a great deal since then. First, the basic set of keyboard characters, the so-called ASCII set, was extended, so that the commoner non-English diacritics could be included. But even then it only allowed up to 256 characters – and there are far more letter shapes in the world than that. Just think of the array of shapes you find in Arabic, Hindi, Chinese, Korean, and the many other languages which don't use the Latin alphabet. Today, a new coding system, the UNICODE system, is much more sophisticated: in its latest version it allows the representation on screen of over 94,000 characters - but that's still well short of the total number of written characters in all the world's languages, which have been estimated to be c.175,000.

My feeling is that the future looks good for Web multilingualism, and a number of influential people seem to share this view. Ned Thomas, for instance, is editor of a bulletin called *Contact* – the quarterly publication of the European Bureau of Lesser Used Languages. In an editorial in 2000 he said: 'It is not the case ... that all languages will be marginalized on the Net by English. On the contrary, there will be a great demand for multilingual Web sites, for multilingual data retrieval, for machine translation, for voice recognition systems to be multilingual.' And Tyler Chambers, the creator of various Web language projects, agrees: 'the future of the Internet', he says, 'is even more multilingualism and cross-cultural exploration and understanding than we've already seen.' I concur. The Web offers a World Wide Welcome for global linguistic diversity. And in an era when so many languages of the world are dying (estimates suggest that half the world's languages will have disappeared by the end of this century), such optimism is truly revolutionary.

A new medium for language doesn't turn up very often, which is why the linguistic effects of electronic communications technology have attracted so much attention. And with mobile phones, where the small-screen technology is so constraining, the effects have generated one of the most idiosyncratic varieties in the history of language. I call it Textspeak.

Its chief characteristic is rebus abbreviation - forms made out of a combination of letters, letters representing syllables, and logograms (such as & and numerals), as seen in *NE1*, *2day*, *B4*, and *l8r* ('later'). Such forms are by no means restricted to Textspeak; they turn up in other electronic domains, such as e-mails and chatgroups. Indeed, rebuses have

a much longer linguistic history. The Victorians played games with them, and children's Christmas annuals have long contained puzzles using them. But in Textspeak something more radical has taken place.

The nature of telephony, plus the on-screen limitation to 160 characters, has motivated a much more wide-ranging and innovative set of conventions. Textspeak has its own range of direct-address items, such as *F2T* ('free to talk?'), *PCM* ('please call me'), *MMYT* ('Mail me your thoughts'), and *RUOK* ('are you OK?'). Multi-word sentences and response sequences can be used, reduced to a sequence of initial letters. Typical examples are *SWDYT* ('So what do you think?'), *BCBC* ('Beggars can't be choosers'), *BTDT* ('Been there, done that'), *YYSSWW* ('Yeah, yeah, sure, sure, whatever'), and *HHOJ* ('Ha, ha, only joking'). Even more ingenious coded abbreviations have been devised, especially among those for whom argot is a desirable safeguard against unwelcome surveillance.

Users seem to be aware of the high information value of consonants as opposed to vowels, judging by such vowel-less items as *XLNT* ['excellent']. And there is ergonomic value in abbreviation, too, given that the number of key-strokes saved bears a direct relationship to time and energy - and (depending on your service-provider) maybe even the eventual size of your telephone bill. In a creation such as *ru2cnmel8r* ('Are you two seeing me later?'), the full form uses over twice as many key-strokes.

In 2002 I compiled a text-messaging dictionary, and - ignoring the difference between upper-case and lower-case usage - collected about 500 Textspeak abbreviations. However, only a small number of these actually turn out to be in regular use. The vast majority are there just to be 'clever', illustrating the possibilities of language play. *ROTFL* ('rolling on the floor laughing') may have had some use at the outset, but its later developments (such as *ROTFLMAO* and *ROTFLMAOWTIME* - 'rolling on the floor laughing my ass off ... with tears in my eyes') illustrate idiosyncrtatic communicative one-upmanship rather than genuine community usage. And I doubt whether many texters actually use such creations as *LSHMBB* ('laughing so hard my belly is bouncing').

The method isn't without its difficulties. Leaving out letters always runs the risk of ambiguity. From the receiver's point of view, a single sequence can have more than one meaning: *BN* - 'been' or 'being', *CID* - 'consider it done' or 'crying in disgrace', *CYA* - 'see you' or 'cover your ass', *N* - 'and' or 'no', *Y* - 'why' or 'yes'. If a message of transmitted love gets the reply *LOL*, it is up to you to decide whether the response was 'laughing out loud' or 'lots of love'. And you have to know your recipient before you decode *GBH*, which can be either a 'great big hug' or 'grievous bodily harm'. There are similar ambiguities in the Textspeak of other langauges.

From the sender's point of view, there are also choices to be made. 'Good to see you' can be *GTCY*, *GTSY*, *G2CY*, or *G2SY*; 'I love you' can be *ILU*, *ILUVY*, or *ILY*; 'thanks' can be *THNX*, *THX*, *TX*, or *TNX*. I found a remarkable eight variants for 'talk to you later': *TTUL*, *TTUL8R*, *TTYL*, *TTYL8R*, *T2UL*, *T2UL8R*, *T2YL*, and *T2YL8R*, and there are probably others Even more exist for 'what's up?' - depending on how many U's you bother to send: *WASSUP*, *SUP?*, *WU?*, *WSU?*, *WSUU?*, *WSUUU?*, etc. Doubtless text-messaging dialects are already evolving.

What is not clear is just how limiting this technology is, as a messaging system. There must be a serious limit to the amount of information which can be conveyed using abbreviation, and a real risk of ambiguity as soon as people try to go beyond a stock set of social phrases. The set of possible messsages is really very small, and only a few abbreviations - such as *C* ('see'), *B* ('be'), *2* ('to, too, two'), *4* ('for, four, -fore'), and *U* ('you') - can be used in lots of sentences.

The constraints will become increasingly apparent as people try to adapt the technology to grander designs, such as Internet access. While it is possible in principle to download Internet pages onto our mobile phone screen or the display of our personal digital assistant, what do we lose, informationally speaking, when a graphically elaborate text is reduced to such a scale? What kind of linguistic 'translation' needs to take place in order to ensure that the sentence structures used on the small screen are manageable and intelligible? It seems inevitable that sentence length will tend to be short, and that complex sentence structure will be avoided.

Will Textspeak have an effect on the language as a whole? This is unlikely. The whole point of the style is to suit a particular technology where space is at a premium; and when that constraint is dropped, abbreviated language no longer has any purpose. Its 'cool' associations amongst young (or at least, young-minded) people will allow some of its idiosyncrasy to achieve a use elsewhere, and there are occasional reports of Textspeak creeping into other forms of writing, such as school essays. But these are minor trends, part of the novelty of the medium. They can be controlled as part of the task of developing in children a sense of linguistic appropriateness - in the UK, one of the basic principles behind the National Curriculum in English.

Some people object to Textspeak. Some are bemused by it. I am fascinated by it, for it is the latest manifestation of the human ability to be linguistically creative and to adapt language to suit the demands of diverse settings. In Textspeak, we are seeing, in a small way, language in evolution.