

The Language Revolution

DAVID CRYSTAL

Not like speech

Computer-mediated communication is not like speech, even in those electronic situations which are most speech-like, such as e-mailing. There are several major differences between CMC and face-to-face conversation. The first is a function of the technology – the lack of simultaneous feedback. The success of a conversation totally depends on the participants providing each other with immediate feedback. While you speak to me, I do not stay unmoved and silent: my face and voice provide an ongoing commentary on what you are saying. Nods and smiles work along with a wide range of vocalizations, such as *uh-huh*, *yeah*, *sure* and *ooh*. Such messages from our listeners tell us how we are doing, and we react to them instinctively and immediately. A puzzled look makes us rephrase.

polity

A doubtful *hmm* makes us rethink. Without these, a conversation quickly breaks down, or becomes extremely stilted and artificial. It is difficult enough on the telephone, when just the visual cues are absent. Imagine the difficulty in a face-to-face conversation if both visual and auditory feedback were missing.

But that is how it is in e-mail and chatgroup interaction. Messages sent via a computer are complete and undirectional. When we send a message to someone, we type it a keystroke at a time, but it does not arrive on that person's screen a keystroke at a time – in the manner of the old teleprinters. The message does not leave our computer until we 'send' it, and that means the whole of a message is transmitted as a unit, and arrives on the recipient's screen as a unit. There is no way that a recipient can react to our message while it is being typed, for the obvious reason that recipients do not know they are getting any messages at all until the text arrives. Correspondingly, there is no way for a sender to get a sense of how successful a message is while it is being written – whether it has been understood, or whether it needs repair. There is no technical way (at the moment) which would allow the receiver to send the electronic equivalent of a simultaneous nod, an *uh-uh*, or any of the other audio-visual reactions which play such a critical role in face-to-face interaction. Messages cannot overlap. As a result, recipients are committed to experiencing a waiting period before the text appears – on their screen there is nothing, and then there is something, an 'off-on' system which well suits the binary computer world but which is far removed from the complex realities of everyday conversation. This factor alone makes e-conversations totally unlike those which take place in 'real world' speech.

The second big difference between Netspeak and face-to-face conversation can be illustrated from real-time chat-

rooms. If you are in a chatroom, talking around a particular theme, you see on your screen messages coming in from all over the world. If there are thirty people in the room, then you could be seeing up to thirty different messages, all making various contributions to the theme, but often clustering into half a dozen or so sub-conversations. It is a bit 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 all of them or contribute to all of them at the same time. In a chatroom you can't avoid attending to them, and you can contribute to as many as your mental powers and typing speed permit. It has never been possible before, in the history of human communications, to participate simultaneously in multiple conversations. Now you can. This too is a revolutionary state of affairs, as far as speech is concerned.

A third difference results from the temporal limitations of the technology: the rhythm of an Internet interaction is very much slower than that found in a speech situation, and disallows some of conversation's most salient properties. With e-mails and asynchronous chatgroups, a response to a stimulus may be anything from seconds to months, the rhythm of the exchange very much depending on such factors as the recipient's computer (e.g. whether it announces the instant arrival of a message), the user's personality and habits (e.g. whether messages are replied to at regular times or randomly) and the circumstances of the interlocutors (e.g. their computer access). The time-delay (usually referred to as *lag*) is a central factor in many situations: there is an inherent uncertainty in knowing the length of the gap between the moment of posting a message and the moment of receiving a reaction. Because of lag, the rhythm of an interaction – even in the fastest Netspeak encounters – lacks the pace and predictability of that found in telephonic or face-to-face conversation.

Even if a participant types a reply immediately, there may be a delay before that message reaches the other members' screens, due to several factors, such as bandwidth processing problems, traffic density on the host computer, or some problem in the sender's or receiver's equipment.

Lag problems make chatroom interactions very different from anything humans have experienced in dialogue before. The frustration is on both sides of the communication chain. From the sender's point of view, the right moment to speak may be missed, as the point to which the intended contribution related may have scrolled off the screen and be fast receding from the group's communal memory. And from the recipient's point of view, the lack of an expected reaction is ambiguous, as there is no way of knowing whether the delay is due to transmission reasons or to some 'attitude' on the sender's part. Unexpected silence in a telephone conversation carries a similar ambiguity, but at least there we have well-established turn-taking manoeuvres which can bring immediate clarification ('Hello?', 'Are you still there?'). The linguistic strategies which express our face-to-face conversational exchanges are much less reliable in chatgroups. Colin may never get a reaction to his reply to Jane because Jane may never have received it (for technical reasons), may not have noticed it (because there are so many other remarks coming in at the same time), may have been distracted by some other conversation (real or on-line), may not have been present at her terminal to see the message (for all kinds of reasons), or simply decided not to respond. Equally, she may have replied, and it is *her* message which has got delayed or lost. When responses are disrupted by delays, there is little anyone can do to sort such things out.

The larger the number of participants involved in an interaction, the worse the situation becomes. Delays in a

conversation between two people are annoying and ambiguous, but the level of disruption is usually manageable, because each person has only one interlocutor to worry about. If a simple e-mail situation is affected by serious delay, feedback via phone or fax is easily providable. But when an electronic interaction involves several people, such as in chatgroups, virtual worlds and e-mails which are copied repeatedly, lag produces a very different situation, because it interferes with another core feature of traditional face-to-face interaction: the conversational *turn*. Turn-taking is so fundamental to conversation that most people are not conscious of its significance as a means of enabling interactions to be successful. But it is a conversational fact of life that people follow the routine of taking turns when they talk, and avoid talking at once or interrupting each other randomly or excessively. Moreover, they expect certain 'adjacency-pairs' to take place: questions to be followed by answers, and not the other way round; similarly, a piece of information to be followed by an acknowledgement, or a complaint to be followed by an excuse or apology. These elementary strategies, learned at a very early age, provide a normal conversation with its skeleton.

When there are long lags, the conversational situation becomes so unusual that its ability to cope with a topic can be destroyed. This is because the turn-taking, as seen on a screen, is dictated by the software, and not by the participants: in a chatgroup, for instance, even if we did start to send a reaction to someone else's utterance before it was finished, the reaction would take its turn in a non-overlapping series of utterances on the screen, dependent only on the point at which the send signal was received at the host server. Messages are posted to a receiver's screen linearly, in the order in which they are received by the system. In a multi-user environment, messages are coming in from various sources all the time, and with different lags.

Because of the way packets of information are sent electronically through different global routes between sender and receiver, it is even possible for turn-taking reversals to take place, and all kinds of unpredictable overlaps. The time-frames of the participants do not coincide. Lucy asks a question; Sue receives it and sends an answer; but on Ben's screen the answer is received before the question. Or, Lucy sends a question, Sue replies, and Lucy sends another question; but on Ben's screen the second question arrives before Sue's reply to the first. The situation may be further complicated if Sue (or anyone) decides to answer two questions from different participants, sending them together. Steve, meanwhile, copied in to the exchanges, is out of the office, and responds a day later, after other messages have come in. There are enormous possibilities for confusion once orderly turn-taking is so disruptable and adjacency-pairs are so interruptible. What is surprising is that practised participants seem to tolerate (indeed revel in) the anarchy which ensues.

Issues of feedback and turn-taking are ways in which computer-mediated interaction differs from conversational speech. But it is unlike speech also with respect to the formal properties of the medium – properties that are so basic that it becomes extremely difficult for people to live up to the recommendation that they should 'write as they talk'. Chief among these properties is the domain of *tone of voice* – 'it ain't what you say but the way that you say it' – as expressed through vocal variations in pitch (intonation), loudness (stress), speed, rhythm, pause and other vocal effects. There have been somewhat desperate efforts to replace tone of voice on screen in the form of an exaggerated use of spelling and punctuation, and the use of capitals, spacing and special symbols for emphasis. Examples include repeated letters (*aaaaahhhhh, soooo*), repeated punctuation marks (*whohe???, hey!!!*), and con-

ventions for expressing emphasis, such as *the *real* point*. These features are capable of a certain expressiveness, but the range of meanings they signal are few, and restricted to gross notions such as extra emphasis, surprise and puzzlement. Less exaggerated nuances are not capable of being handled in this way.

Related to this is the way Netspeak lacks the facial expressions, gestures and conventions of body posture which are so critical in expressing personal opinions and attitudes and in moderating social relationships. The limitation was noted early in the development of Netspeak, and led to the introduction of *smileys* or *emoticons* – combinations of keyboard characters designed to show an emotional facial expression. The two basic types express positive attitudes and negative attitudes respectively (the omission of the 'nose' element seems to be solely a function of typing speed or personal taste):

:-) or :) :-(or :(

Hundreds of ludic shapes and sequences have been invented and collected in smiley dictionaries, some extremely ingenious and artistic, but hardly any are used in serious communication. It is plain that they are a potentially helpful but very crude way of capturing some of the basic features of facial expression. They can forestall a gross misperception of a speaker's intent, but an individual smiley still allows a huge number of readings (happiness, joke, sympathy, good mood, delight, amusement, etc.) which can be disambiguated only by referring to the verbal context. Without care, moreover, they can lead to their own misunderstanding: adding a smile to an utterance which is plainly angry can increase rather than decrease the force of the 'flame'. It is a common experience that a smile can go down the wrong way.

The fact that smileys turn up at all in e-mails and chat-room interaction is indicative of the un-speech-like nature of the medium which the participants are using. Smileys evolved as a way of avoiding the ambiguities and misperceptions which come when written language is made to carry the burden of speech. They are brave efforts, but on the whole Netspeak lacks any true ability to signal facial meaning, and this, along with the unavailability of tones of voice, places it at a considerable remove from spoken language. One day, developments in interactive technology will allow us to see and hear other participants while they are talking, so that some of these limitations will be eliminated; but there will always be properties of the electronic medium which will enable us to use language in ways that traditional speech could never perform.

Not like writing

If Netspeak does not display the properties we would expect of speech, does it instead display the properties we expect of writing? Here too, there are fundamental differences. Let us consider first the space-bound character of traditional writing – the fact that a piece of text is static and permanent on the page. If something is written down, repeated reference to it will be an encounter with an unchanged text. We would be amazed if, upon returning to a particular page, it had altered its graphic character in some way. Putting it like this, we can see immediately that computer-mediated communication is not by any means like conventional writing. A 'page' on the Web often varies from encounter to encounter (and all have the option of varying, even if page-owners choose not to take it) for several possible reasons: its factual content might have

been updated, its advertising sponsor might have changed, or its graphic designer might have added new features. Nor is the writing that you see necessarily static, given the technical options available which allow text to move around the screen, disappear/reappear, change colour, and so on. And from a user point of view, there are opportunities to 'interfere' with the text in all kinds of ways that are not possible in traditional writing. A page, once downloaded to the user's screen, may have its text cut, added to, revised, annotated, even totally restructured, in ways that make the result seem to come from the same source as the original. The possibilities are causing not a little anxiety among those concerned about issues of ownership, copyright and forgery.

The other Internet situations also display differences from traditional writing, with respect to their space-bound presence. E-mails are in principle static and permanent, but routine textual deletion is expected procedure (it is a prominent option in the management system), and it is possible to alter messages electronically with an ease and undetectability which is not possible when people try to alter a traditionally written text. What is especially revolutionary about e-mails is the way the medium permits what is called *framing*. We receive a message from M which contains, say, three different points in a single paragraph. We can, if we 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 we send back to M then looks a bit like a play dialogue. Then, M can do the same thing to our responses, and when we get the message back, we see M's replies to what we sent. We 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 has never been anything like this in the history of human written communication. Although in principle it might have been possible to receive a letter, cut it up into strips, intercalate our responses, then paste everything onto another sheet of paper before returning it to its sender, this would hardly count as normal behaviour. But we do this all the time with e-mails, without thinking twice (once we are used to it).

Other features of computer-mediated communication take us even further away from traditional writing. Probably the most important is the *hypertext* link – the jump that users can make if they want to move from one page or site to another. The hypertext link is the most fundamental functional property of the Web, without which the medium would not exist. It does have parallels in some of the conventions of traditional written text. The use of note indicators is a sort of primitive hypertext link, moving the eye from one part of a page to another, or from one page of a text to another (if the notes are collected at the back of a book, as in the present volume, for example). The use within a sentence of bibliographical citations or cross-references (such as 'see p. 333') is another opportunity for a reader to break away from the conventions of linear viewing. But these features are marginal to traditional written language; we can easily think of texts which have no notes or cross-reference citations at all. The Web, by contrast, could not exist without its hypertext links. As Tim Berners-Lee once put it: 'Free speech in hypertext implies the "right to link", which is the very basic building unit for the whole Web.'⁶ There is nothing in traditional written language which remotely resembles the dynamic flexibility and centrality of Web hypertext linkage.

A few other characteristics of traditional written language also display an uncertain relationship to Netspeak, but these hardly fall into the category of 'revolutionary'.

E-mails and chatgroup interactions, where the pressure is strong to communicate rapidly, lack the carefully planned, elaborate construction which is characteristic of so much writing. At one extreme, it might well appear that a revolution is taking place. Some people are happy to send messages with no revision at all, not caring if typing errors, erratic capitalization, lack of punctuation and other anomalies are included. This is actually 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 M in which he mis-spells a word, I do not conclude from this that 'M can't spell'. I simply conclude that M is not a good typist or 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. And it is, in any case, not a universal behaviour. There are many e-mailers who take as many pains to revise their messages as they would in non-Internet settings.

On the whole, Netspeak is better seen as written language which has been pulled some way in the direction of speech than as spoken language which has been written down. However, expressing the question in terms of the traditional dichotomy is itself misleading. Computer-mediated communication is not identical to either speech or writing, but selectively and adaptively displays properties of both. It also does things which neither of the other mediums does, presenting us with novel problems of information management. Consider, as an example, the *persistence* of a conversational message in a chatroom – the fact that it stays on the screen for a period of time (before the arrival of other messages replace it or make it scroll out of sight).

This certainly introduces novel properties to the interaction which are not available in speech. It means, for example, that someone who enters a conversation a couple of turns after an utterance has been made can still see the utterance, reflect upon it, and react to it; the persistence is relatively short-lived, however, compared with that routinely encountered in traditional writing. It also means, for those systems that provide an archiving log of all messages, in the order in which they were received by the server, that it is possible in principle to browse a past conversation, or search for a particular topic, in ways that spontaneous (unrecorded) conversation or traditional book-indexing does not permit.

Netspeak is more than an aggregate of spoken and written features. Because it does things that neither of these other mediums does, it has to be seen as a new species of communication. It is more than just a hybrid of speech and writing, or the result of contact between two long-standing mediums. Electronic texts, of whatever kind, are simply not the same as other kinds of texts. They display fluidity, simultaneity (being available on an indefinite number of machines) and non-degradability in copying; they transcend the traditional limitations on textual dissemination; and they have permeable boundaries (because of the way one text may be integrated within others or display links to others). Several of these properties have consequences for language, and these combine with those associated with speech and writing to make Netspeak a genuine 'new medium'.

The consequences of a new medium: within a language

The linguistic effects of the arrival of a new medium of communication are twofold: it initiates change in the

formal character of the languages which use it; and it offers new opportunities for languages to use it. Of the two, it is the first which has attracted all the publicity with respect to the kind of language encountered on the Internet and in related technology, such as mobile phones (cellphones). The apparent lack of respect for the traditional rules of the written language has horrified some observers, who see in the development an ominous sign of deterioration in standards. Text-messaging is often cited as a particular problem. Children of the future will no longer be able to spell, it is said. However, the fact that youngsters abbreviate words in text-messaging using rebus techniques (*b4*, *CUI8er*), initialisms (*afaiik* 'as far as I know', *imho* 'in my humble opinion') or respelling (*/hx* 'thanks') is hardly new or fundamental. People have been using initialisms for generations (*tjfh*, *asafp*, *fvi*), and rebus games have long been found in word-puzzle books. Even the fullest lists of text-message abbreviations contain no more than a few hundred forms, and few of those are commonly used. And because they were devised to meet the needs of economical messaging on a screen which has a limit of 160 characters, there is little motivation to use these deviant forms elsewhere. They lose their 'cool', group-identifying function when they are taken away from the technology, whether mobile phone or computer. The fact that a few kids might start using their abbreviations in places where they have no purpose – such as school essays – is something to be watched, of course. But that is what teaching needs to do. It has long been a principle of modern language teaching – whether foreign or mother-tongue – to inculcate in children a sense of linguistic responsibility and appropriateness. And children need to be taught, if they have failed to develop the intuition for themselves, that text-messaging abbreviations perform a useful function where space is tight and speed is critical, but not elsewhere.

The same point applies to the variations in the writing system that adults as well as children introduce into their e-mails. Many people use a severely reduced system, with virtually no typographic contrastivity at all. There are three main features. The status of *capitalization* varies greatly. Because most of the Internet is not case-sensitive, a random use of capitals has developed, or the use of no capitals at all. There is a strong tendency to use lower-case everywhere. The 'save a keystroke' principle is widely found in e-mails, chatgroups and virtual worlds, where whole sentences can be produced without capital letters marking the beginning of the sentence or proper names. *Punctuation* also tends to be minimalist, and is completely absent in some e-mails and chat exchanges. Here too, a lot depends on personality: some e-mailers are scrupulous about maintaining traditional punctuation; others use it when they have to, to avoid ambiguity; and some do not use it at all, either as a consequence of typing speed, or through not realizing that ambiguity can be one of the consequences. And *spelling* practice is also distinctive. In English, US spelling is more common than British, partly for historical reasons (the origins of the Internet), and partly for reasons of economy, most US spellings being a character shorter than British ones (*color* vs *colour*, *fetus* vs *foetus*, etc.). Nonstandard spellings, heavily penalized in traditional writing (at least, since the eighteenth century), is used without sanction in conversational settings. As already suggested, spelling errors in an e-mail would be assumed to be not an indication of lack of education (though they may be) but a function of typing inaccuracy.

I interpret these developments as extending the expressive range of the languages that are now on-line. Novel conventions are being introduced. This has always happened when a revolutionary communication technology has arrived. When printing was introduced, a whole

range of new manifestations of written language emerged, including innovative layout conventions and punctuation usages, and a gradual standardization of spelling. When telephones arrived, new conventions of discourse interaction had to be devised (saying hello, specifying or confirming the number, etc.). When broadcasting arrived, spoken language diversified immensely, with results which today range from weather-forecasting style to sports commentary. And so it is now with the Internet, where the technology from the outset motivated new kinds of expression. For example, the lower-case default mentality means that any use of capitalization is a strongly marked form of communication. Messages wholly in capitals are considered to be 'shouting', and usually avoided; words in capitals add extra emphasis (with asterisks and spacing also available):

This is a VERY important point.

This is a * very * important point.

This is a v e r y important point.

Another distinctive feature of Internet graphology is the way two capitals are used – one initial, one medial – a phenomenon sometimes called *bicapitalization* (*BiCaps*), seen in such names as *AltaVista*, *PeaceNet* and *CompuServe*, or, more complexly, with *QuarkXPress*. There is an increased use of symbols not normally part of the traditional punctuation system, such as the #. Unusual combinations of punctuation marks can occur, such as (to express pause) ellipsis dots (...) in any number, repeated hyphens (---) or the repeated use of commas (,,). Emphasis and attitude can result in exaggerated or random use of punctuation, such as !!!!!!! or £\$%&%.¹

New vocabulary has also come into languages which are on-line, much of it deriving from the global use of English.

A large number of words and phrases have emerged which are needed to talk about Internet-restricted situations, operations, activities and personnel, making this one of the most creative lexical domains in contemporary English, and analogous developments are taking place in other languages, as they increase their on-line presence. Many terms are associated with the software which enables people to use the Internet, and which routinely appear on screen. Some have a permanent presence (albeit in hidden menus), in the form of the labels used to designate screen areas and functions, and to specify user options and commands: *file, edit, view, insert, paste, format, tools, window, help, search, refresh, address, history, stop, contact, top, back, forward, home, send, save, open, close, select, toolbar, fonts, options*. Some terms appear only at intervals on a screen, depending on circumstances – usually, when things are going wrong, in the form of error messages (there seem to be no positive messages to tell us that everything is going right): *forbidden, illegal operation, error, not found, 404 error* [‘a page or site is no longer in service’]. Several terms are associated with the use of computer hardware: *freeze, lock, down, hang, crash, bomb, client* (the machine, not the user). And terms have emerged for the population of Internet users themselves: *netizens, netters, netties, netheads, cybersurfers, nerds, newbies, surfers, diggerati*. Many of these words are everyday terms which have been given a fresh sense in an Internet context.

A popular method of creating Internet neologisms is to combine two separate words to make a new word, or compound. Some elements turn up repeatedly – for example, *click* in *click-and-buy, one-click, cost-per-click, double-click, clickthrough rate*, and so on. Similar in function are the use of *cyber-* and *hyper-* in such words as *cyberspace, cyberculture, cyberlawyer, cybersex, cybersquatter, cyberian, cyber rights; hypertext, hyperlink, hyperfiction* and *hyperzine*.

Blends (in which part of one word is joined to part of another) are illustrated by *netiquette, infonet, datagram, infobahn* and *Internaut*. An innovation is the retaining of the period found in electronic addresses within certain compounds, as a kind of infix, seen in *net.legend, net.abuse* and *net.citizen*, or sites beginning with *alt.* (with the punctuation mark often spoken aloud as ‘dot’). Acronyms are very common: a tiny sample would include *BBS* [‘bulletin board system’], *BCC* [‘blind carbon copy’], *DNS* [‘domain name system’], *FAQ* [‘frequently asked questions’], *HTML* [‘hypertext markup language’], *ISP* [‘Internet Service Provider’], *URL* [‘uniform resource locator’], and the names of many firms and sites, such as *AOL, IBM* and *IRC*. Letter-plus-number combinations are also found: *W3C* [‘World Wide Web Consortium’], *P3P* [‘Platform for Privacy Preferences’] and *Go2Net*.

It is always a sure sign that a new variety has ‘arrived’ when people in other linguistic situations start alluding to it in speech. It is therefore of considerable interest to note the way in which features of Netspeak have already begun to be used outside of the situation of computer-mediated communication, even though the medium has become available to most people only in the past decade or so. In everyday conversation, terms from the underlying computer technology are given a new application among people who want their talk to have a cool cutting-edge. Examples from recent overheard conversations include *It’s my turn to download now* (‘I’ve heard all your gossip, now hear mine’), *I need more bandwidth to handle that point* (‘I can’t take it all in at once’), *Get with the programme* (‘Keep up’) and *She’s multi-tasking* (‘She’s doing several things at once’). Resonances are now routine on radio and television: presenters commonly add e-addresses when telling listeners and viewers how they might contact a

programme, using *at*, *dot* and *forward slash* to punctuate their utterance. *Dot com* is now a commonly heard phrase, as well as appearing ubiquitously in writing in all kinds of advertising and promotional material. And the *e-* prefix is now encountered in thousands of expressions, such as *e-text*, *e-zine*, *e-cash*, *e-government*, *e-bandwagon*, *e-books*, *e-conferences* and *e-voting*. In 1998 the American Dialect Society named *e-* 'Word [*sic*] of the Year' as well as 'Most Useful and Most Likely to Succeed'.

How many of these developments will become a permanent feature of English in the twenty-first century it is impossible to say. The same point applies to the impact of the Internet on other languages. We can never predict language change, only recognize it once it has happened. But it is already evident that, since the 1990s, a notion of Netspeak has begun to evolve which is rapidly becoming a part of popular linguistic consciousness, and evoking strong language attitudes. This is bound to grow as the century proceeds.

The consequences of a new medium: for all languages

There is 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 is a story that the 8-year-old son of Kyrgyzstan's President Akayev 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 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.'⁷ This view is no longer valid. With the Internet's globalization, the presence of other languages has steadily risen. By 1998, a widely quoted figure was that about 80 per cent of the Net was in English. This figure derived from the first major study of language distribution on the Internet, carried out in the previous year 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. 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 AltaVista was predicting that by the end of 2002 less than 50 per cent of the Web would be in English. This has turned out to be the case. In certain parts of the world, the local language is already dominant. According to one Japanese Internet author, Yoshi Mikami, 90 per cent of Web pages in Japan are already in Japanese.

It is not surprising to see the Web (and the Internet as a whole) becoming predominantly *non-English* as communications infrastructure develops in Asia, Africa and South America. This is where the bulk of the people are. The Web is increasingly reflecting the distribution of language presence in the real world, and many sites provide the

evidence. There are thousands of businesses now doing their best to present a multilingual identity, and hundreds of major sites collecting all kinds of data on the languages themselves. Call up the font archive at the University of Oregon, for example: you'll find 112 printing fonts in their archives for over forty languages. They have a nice sense of humour too – because you'll also find some data there on alien languages, such as Klingon, and folklore languages, such as Elvish, which Tolkien invented for *The Lord of the Rings*. Spend an hour hunting for languages on the World Wide Web and you will 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 listed products for 728 languages. I found an African resource list which covered several local languages: Yoruba, for example, was illustrated by some 5,000 words, along with proverbs, naming patterns and greetings. Another site dealt with no less than eighty-seven 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 1,000 quite quickly. It is not difficult to find evidence of an Internet presence for all the more frequently used languages in the world, and for a large number of minority languages too. My guess is that at least a quarter of the world's languages – about 1,500 – have some sort of cyber-existence now. And this is language presence in a real sense. These are not sites which only analyse or talk about languages, from the point of view of linguistics or some other academic subject. They are 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 are represented by only a sprinkling of sites.

The Internet is the ideal medium for minority languages, and this is a lifeline which could prove to be important for some of the languages whose plight was described in chapter 2. If you are a speaker or supporter of an endangered language – an aboriginal language, say, or one of the Celtic languages – you are 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. The Web message stays around, too, in a way that newspaper and broadcasting references do not. 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. Several of the world's 'smaller' languages that have access to Internet technology – such as the minority languages of Europe and many North American Indian languages – now have Web sites and foster virtual speech communities.

On the other hand, I have to recognize that developing a significant cyber-presence for a language is not easy. To begin with, the infrastructure has to be there – and with so many endangered languages existing in parts of the world where the electricity supply is unreliable or non-existent, the priorities are clear. Then, to use present-day Web technology, the language has to be written down, and as we have seen in chapter 2, this excludes some 2,000

languages which have not been documented at all. Another complication is that the distinctive letters of some languages (especially those which make use of a range of accent marks) are often not easily encodable so that they can be routinely 'read' by computers everywhere. Lastly, even if there is the technology and the literacy, there is a hurdle of motivation to be crossed. There seems to be a sort of 'critical mass' of Internet penetration which has to build up in a country or community before a language develops a vibrant cyber-life. It is 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 is using them and adding to them and talking about them. That is 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 is still evolving, and likely to be one of the main points of development in the next few years. Everything depends on how quickly new sites can build up a local language momentum. And we must not underestimate the practical difficulties. Take, for example, the apparently simple issue of representing a language's letters accurately. 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 symbols 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 there have been important develop-

ments. First, the basic set of keyboard characters, the so-called ASCII set, was extended, so that the commoner non-English accents could be included. Even then it only allowed up to 256 characters – and there are far more letter or word 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 do not use the Roman alphabet. Today, a new coding, the UNICODE system, is much more sophisticated: in its latest version it allows the representation on screen of over 94,000 characters – though that is still well short of the total number of written characters in all the world's languages, which has been estimated to be about 175,000.

My feeling is that the future looks good for Web multilingualism, and this opinion seems to be becoming widespread. 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.'⁸ 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.'⁹ 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, such optimism is truly revolutionary.