David Crystal explains how the internet is turning out to be a friend to all the world's languages

Weaving a Web of linguistic diversity

The global English debate

The World Wide Web is an eclectic medium, holding a mirror up to our linguistic nature. Not only does it offer a home to all linguistic styles within a language; it offers a home to all languages — once their communities have a functioning computer technology. And its increasingly multilingual character has been the most notable change since it started out as a totally English medium.

For many people the language of the internet *is* English. "World, Wide, Web: three English words" was the headline of a piece by Michael Specter in the New York Times a few years ago. The article went on to comment: "If you want to take full advantage of the internet there is only one real way to do it: learn English."

Specter did acknowledge the arrival of other languages: "As the Web grows, 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. But it won't necessarily happen soon."

The evidence is growing that this conclusion was wrong. With the internet's globalisation the presence of other languages has steadily risen. By the mid-90s a widely quoted figure was that about 80% of the Net was in English — a figure supported by the first big 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 — notably German, Japanese, French and Spanish — entering the ring.

Since then the estimates for English have been falling, with some commentators 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 Global Reach survey has estimated that people with internet access in non-English-speaking countries increased from 7m to 136m between 1995 and 2000. In 1998 the total number of new non-English websites passed the number of new English websites.

At a conference on search engine strategies last April, Alta Vista was predicting that by next year less than half of the Web would be in English. English-language author David Graddol has predicted an even lower figure in due course, 40%. In parts of the world the local language is already dominant. According to the Japanese internet author Yoshi Mikami, 90% of Web pages in Japan are now in Japanese.

The Web is increasingly reflecting the distribution of language presence in the real world, and many sites provide the evidence. They range from individual businesses doing their best to present a multilingual identity to big sites collecting data on many languages. Under the first heading we encounter such newspapers as the Belgian daily Le Soir, which is represented by six languages — French, Dutch, English, German, Italian and Spanish. Under the latter heading we find such sites as the University of Oregon Font Archive, providing 112 fonts in its archives for more than 40 languages.

A World Language Resources site lists products for 728 languages. An African resource list covers several local languages; Yoruba, for example,
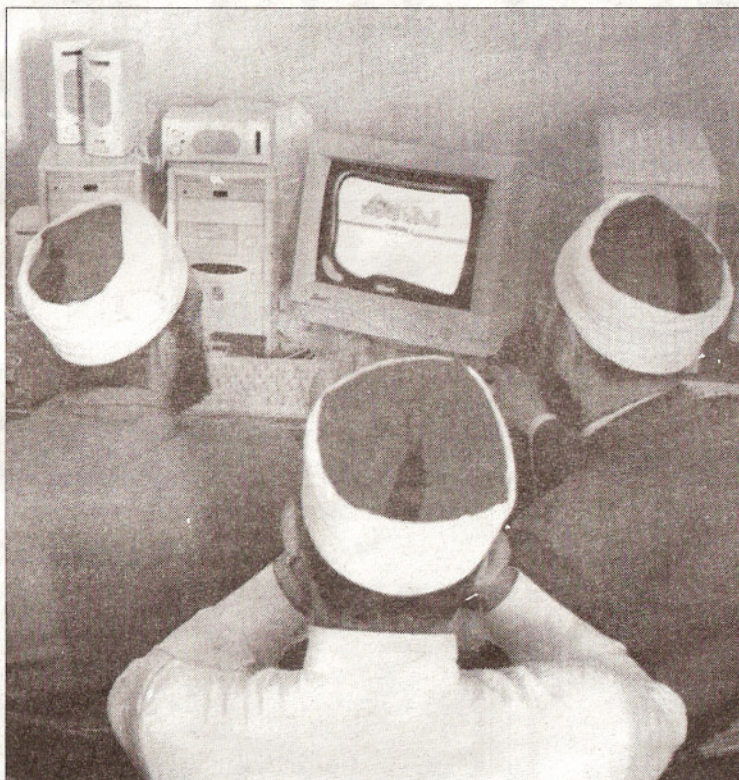
Students at Cairo's al-Azhar University join the small but growing number of Arabic speakers using the internet Photograph: Mohammed Al-Sehiti

is illustrated by some 5,000 words, along with proverbs, naming patterns and greetings. Another site deals with 87 European minority languages. Some sites are small in content, but extensive in range: one gives 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 have found more than 1,000. It is not difficult to find evidence of a Net presence for the vast majority of the more frequently used languages, and for a large number of minority languages too. I would guess that about a quarter of the world's languages have some sort of internet presence.

In all these examples we are encountering language presence in a real sense. These are not sites that only analyse or talk about languages;

they allow us to see languages as they are. In many cases, the total Web presence, in terms of number of pages, is small. The crucial point is that the languages are out there, even if represented by only a sprinkling of sites. It is the ideal medium for minority languages, given the relative cheapness and ease of creating a Web page, compared with the costs of print, TV or radio.

However, developing a significant cyber-presence is not easy. Until a critical mass of internet penetration in a country builds up, and a corresponding mass of content exists in the local language, the motivation to switch from English-language sites will be limited to those for whom issues of identity outweigh issues of information. The future is also dependent on the levels of English-speaking ability in individual countries, and the further growth in those levels.

There are also practical problems, though a great deal has been done since the mid-90s to address them. First, the Ascii character set still fails to adequately support the array of letter shapes in Arabic, Hindi, Chinese, Korean and the many other languages in the world that do not use the Latin alphabet. The Unicode coding system, the alternative to Ascii, allows more than 65,000 characters; but the implementation of this system is still in its infancy. The Web consortium has an internationalisation activity looking specifically at different alphabets, so that operating systems can support a page in any alphabet.

The future looks good for Web multilingualism. As Ned Thomas commented last year in an editorial for Contact, the bulletin of the European Bureau of Lesser Used Languages: "It is not the case . . . that all languages will be marginalised on the Net by English. On the contrary, there will be a great demand for multilingual websites, 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 is even more multilingualism and cross-cultural exploration and understanding than we've already seen."

I agree. The Web offers a World Wide Welcome for global linguistic diversity.

David Crystal is honorary professor of linguistics at the University of Wales, Bangor. His book Language And The Internet will be published by Cambridge University Press later this year