

Information Technology: A Challenge for Minority Languages

THE TRIUMPH OF THE COMPUTER AGE IS A
DOUBLE-EDGED SWORD FOR LESS-SPOKEN TONGUES.

OPINION BY
NICHOLAS OSTLER

Is the development of IT setting up an "appointment with destiny" for all the languages of the world, an opportunity which they ignore at their peril, a last, best chance to establish themselves in the 21st century?

Certainly, the councils of Europe now take it for granted that, in our age, the future of languages depends on their fitness for use in information technology:

... many of the minority languages are experiencing difficulties, often under the influence of changing patterns of communication. Penetration of the new technologies could substantially accelerate this process, threatening to diminish the linguistic and cultural diversity of European society.

... The rapid rise in use of information and communication technologies will naturally favour languages which can be successfully processed. Languages supported by key software products offering powerful facilities for manipulating text also provide almost unlimited access to information services in those languages... The long-term viability of languages not specifically supported is therefore put at risk.

EC proposal for a Council Decision,
Multilingual Information Society, 1995

On the one hand, we can see that in most of the world's dominant languages, documents can be readily input, transmitted, and represented as text held in computers. English sets the standard; but Herculean efforts in the 1970s and 1980s have ensured that even Japanese (with its 3,000 characters and non-phonetic spelling) can now be accommodated comfortably.

At the other end of the spectrum, languages where this is still not easy confront serious problems in taking part in the world's discourse, which is increasingly electronic. The difficulties may be of various kinds—inappropriate keyboards, competing or absent standards for representing the distinctive written form of a language, inadequate spell-checkers, absent electronic dictionaries, error-prone OCR, whatever—but to the extent that they exist, the



Nicholas Ostler

language will be an inconvenient medium for information, and hence will tend to be used less, even by those for whom it should be most convenient—its own native speakers.

Yet in fact, at its basic storage and processing levels, information technology is not about human language at all. And for the programmer, a human language is a rather awkward code in which to store information, neither thorough enough to record all details that may be relevant, nor elegant enough as a concise abbreviation of systematic patterns. As a result, real natural-

language technology, which would go beyond representing the text on the page and try to get at the meaning, has yet to arrive. It could be painted as the Cinderella of information media, with a long-standing invitation to the ball, but unable to contact its fairy godmother.

Compared with music, pictures, or video, language material is still far less integrated in mass-use systems. These other types of material still face problems, e.g., in how to keep up with users' performance expectations yet stay within the current limits of data storage and flow; but there is no difficulty in representing their information in usable form. For language, even in its most studied variety (English), we are only just now beginning to see systems that can convert between its spoken and written forms at a tolerable level of competence. Searching through language data does not go beyond spotting key words: computers still can't use the meaning connections—which are the real point of language—to find collateral information that a text is talking about, let alone draw inferences from what is said to what is not.

PROCESSED NOT USED

The language that we all think in, then—however familiar we may be with doing things on a computer—is not part of information technology. Language is *processed* by IT, rather than *used*. The machines are best seen as glorified notepads, dictionaries, libraries, and postal systems, rather than as secretaries, teachers, librarians, or even postal employees. Certainly not colleagues.

This does not mean that they are not useful, of course. And the power of these new tools goes beyond their real utility—insidiously so. New technology, despite what its promoters say, is never just a useful addition to what we had already: it tends to replace it, and any advantages of the old ways will be lost. We don't get *extra*, we get *different*.

The consolation is that we can hope these losses are more than compensated by the advantages of the new. The vast capacities of memory seen in pre-literate societies all over the world, for laws and ceremonials and literature, are no more exercised, and so are lost to us—but what do we lose when we can look up, and read out, a far wider range of things we never had to memorize? Our, and our children's, powers of mental arithmetic have wilted before the pocket calculator; but this, among many other benefits, has inspired the mathematics of chaos theory. The thrill, or chill, of receiving a telegram personally delivered, has died away before the daily mass of instant email and fax messages from across the world.

What does all this have to say to the plight of minority languages? These languages (Welsh, Basque, Sorbian, Kashubian—the list just in Europe is three times as long as the list of national languages) are repositories of tradition or they are nothing—although it is certainly to be hoped that they are much more than that. Must we suppose that—like the telegram, the memorized multiplication table, and the Icelandic lawspeaker—they must fade away before the searing dawn of large languages equipped with the latest in electronic coding techniques?

BIG BOYS TALK LOUDER

The question is exceedingly difficult to answer. First, because it is about the future; but also because (as we have seen) the aid which IT gives to language use is mostly auxiliary, and does not go to the heart of what a language does: to aid communication and solidarity between its speakers, and to channel and focus their thinking. There is no electronic mode of discourse that is available only to the big boys, the users of the big languages of the developed world. It is just that the big boys can talk louder, and more frequently among themselves.[1]

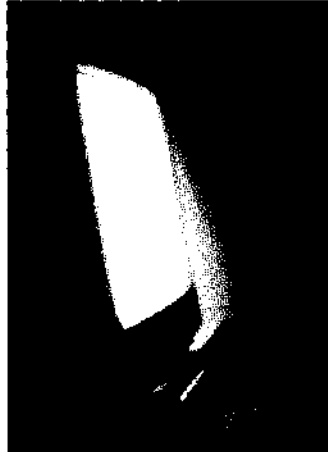
It is difficult for speakers of minority languages, essentially lacking state governments to look after their affairs, to ensure that they are represented in the councils which set coding standards or in the investment to research and implement IT which serves them. But this difficulty can have a positive or a negative effect. It can act as a challenge, a spur to endeavor. Or it can act as a discouragement, an obstacle to be added to all the others which beset minorities. There is no determinism about the future when human response is crucial.

The former effect, fortunately, seems to be easier to see than the latter. The Catalan language, although spoken by a minority within one of the less rich states of the EU, nevertheless leads the field in development of language corpora, dictionaries, and coding standards.[2] The Strategy of the Welsh Language Board[3] (set up by Act of Parliament in 1993) explicitly refers to the role of information technology in support of Welsh. Further afield in Mexico, the CELIAC organization[4] provides an example of how language technology can speed the advent of literacy: speakers of minority languages, hitherto largely unwritten, are encouraged to absorb some basic principles of how to represent a language

graphically, and then to get writing on PCs, and start creating a written literature.

THE VIRTUES OF VIRTUALITY?

Some features of IT may have a distinctive effect on minority languages, but one that cannot yet be discerned clearly. With email and the Internet, it is now much easier to set up virtual communities which are scattered far and wide geographically: will this benefit minority languages?



The answer is rather finely balanced. On the one hand, minority languages all owe their origin to smaller, physical communities; a virtual community implicitly dilutes this, and might weaken the traditional community, when its members are free to look beyond it. But on the other hand, the kind of community benefits shared over the Internet are opposite to those of a small traditional community: it is long on openness and freedom, short on solidarity and warmth. And even when traditional communities find it hard to sustain themselves, the electronic media may enable discourse in their own languages to continue among its members after they have scattered: Scottish, and increasingly Irish and Gaelic may be good examples here, where the center of

gravity of the language is shifting more and more to the cities from the village localities of the Gaeltachtaí. A permanent community, existing over centuries, is probably essential to create a language—it may not be required to sustain it.

Some bizarre effects are already seen on the socio-linguistics of Wales, due not just to the Internet but the general penetration of Welsh into electronic media. As seen in the latest (1991) census figures, the usage of Welsh is down in the traditional areas of the North and West (as monolingual English people penetrate the village communities, looking for a quiet rural life). But amazingly, the percentage of Welsh-speakers throughout the previously English-speaking area around Cardiff and the South has gone up: an urban network community is springing up of Welsh-speakers taking advantage of its privileged status in the media, education, and government.[5]

IRONIES

So two ironies may yet sustain minority languages in the current electronic age. One is that, despite the publicity, the basic function of the technology is not to *process* language but to *disseminate* it: to send and broadcast messages and texts, so as to support communities that were previously beyond the range of one-to-one discourse. Email and the Internet above all disseminate text, but with a speed and convenience that gives a whole new life to correspondence.

The other is that the language used in these communications is all (or mostly) monolingual.[6] For real exchange of ideas and emotions, automatic translation is still quite unfeasible—indeed, it is a good example of the technology of language which, as we noted at the outset, has not been consummated. As a result, language becomes a badge of membership, a *laissez passer*, an earnest of sincere interest to enter these communities. The failure of language technology to make language a transparent medium is serving, at present, to tie language much closer to its culture: you can't have one without the other.

(continued on page 40)♦

NOTES


[1] But the glut of material that this causes is a well-known problem in its own right. (And effective information filtering is still very much a pious aspiration, not an achievement, of language technology.)

[2] <http://www.iec.es/>

[3] http://www.netwales.co.uk/byig/home_eng.htm

[4] Centro Editorial de Literatura Indígena, Asociación Civil: celiac@infosel.net.mx (Spanish); ufruss@nersp.nerdc.ufl.edu (English).

[5] J. Aitchison & H. Carter, *A Geography of the Welsh Language 1961-1991*, Univ. Wales Press 1994.

[6] The widespread availability of parallel pages on the Internet (typically English versions of pages written in other languages) is a limit on this: but for person-to-person communication (e.g., by email) this kind of pre-digested translation is of course unavailable. 

*Nicholas Ostler is a Contributing Editor of Language International and Director of Linguacubun Ltd, a UK language consultancy. He also edits *Iatiku*, the newsletter of the Foundation for Endangered Languages.*