Fighting Words

AS THE WORLD GETS SMALLER, MINORITY LANGUAGES STRUGGLE TO STAKE THEIR CLAIM

BY NICHOLAS OSTLER

Nicholas Ostler

What's in a name? Or in a word, for that matter? Extreme prejudice, at least in some languages, it would seem. Just as in English, "calling names" or "having words with someone" has a decidedly hostile overtone, the very name for "word" in some languages, is derived from "quarrel." That is true of Irish Gaelic, where the normal word for "word" or "verb" is *Briathar*, from the Indo-European **bhrei-trâ* (quarrel): the same source resulted in the Welsh *Brwydr* "a battle." And the neighbors over the water are no more pacific: Scots Gaelic *Bruidhinn*, now the ordinary word for "talk, speak," comes from Old Irish *Bruiden*, which also meant "quarrel" and still does in Modern Irish, *Bruíon*.

Further afield from the British Isles, there is a comparable hullabaloo among the Slavs, where one word for "say," *Molvit'*, appears related to *Mluva*, "tumult," the same root turning up as "speak" or "say" in Persian Avestan *Mru-*, and Indian Sanskrit *Bru-*.

Talk was evidently a truculent business for our ancestors, and one can well understand the motives attributed to the creator goddess Iatiku, working in quite another part of the world, among the Keres peoples of New Mexico; she is held to have made people speak different languages so that it would not be so easy for them to quarrel.

All this goes against a received view in the West, which holds that the fewer languages the better. Surely, a common language will conduce to common understanding.

Unfortunately, recent history suggests on the contrary that Iatiku had it right: violence tends to come when people understand each other all too well. Surrounded by a sea of other languages, Ruanda and Burundi happen to be overwhelmingly monolingual countries, with dialects (Kinyarwanda and Rundi) that are quite intelligible to each other. Croatia, Bosnia, and Yugoslavia (with the exception of Kosovo) are all "divided by a common language" which until the 1990s was often called just *nas jezik*, "our language." And in Southeast Asia, it is notable that the most devastating fighting has come since the war in Vietnam and Cambodia, who alone have official languages spoken by more than 85 percent of the population.

In a much smaller and less deadly way, the fatal presumption in favor of monolingualism can be seen causing strife where there was previously just hubbub. On December 22, 1998, the AOL/UK's *Peace In Ireland* message board carried

A real-life directive from AOL:

"This board is designed for English speakers and Gaelic postings are not allowed. Continuation of this type of posting will result in a warning to the account...posts in Gaelic will be removed without further recourse to the person who posted it."

the following statement from its manager: "This board is designed for English speakers and Gaelic postings are not allowed. Continuation of this type of posting will result in a warning to the account. As for posting in Gaelic and then giving a translation in a later posting, we all know that we then rely on the accuracy and fairness of the person giving the translation. So...to make it crystal clear...posts in Gaelic will be removed without further recourse to the person who posted it."

Predictably, this led to an outcry about censorship and offenses against freedom of speech. About the same time on another list, *Teachers of Celtic Languages* no less, the list owner asked subscribers to "Please show the courtesy to subscribers to this mailing list who do not speak a specific or any Celtic language—include a short summary in English of your posting." One subscriber wrote back at once: "It is sad that it should be regarded as discourteous when people use their own native language. Do you not think that it is inappropriate to force people to communicate in a language which is not their own? I am proud of my language. I have taken great pains to learn to speak and communicate well in it. I use Irish in my work and in my home. Why must I always have to translate? Why all this compulsory English?"

The appeal to courtesy is particularly galling, since it suggests that giving rein to linguistic diversity is somehow self-indulgent, and disrespectful to those unfortunate enough not to know your language. Language learning can be hard, but no one should forget that monolingualism in the majority language is a restricted lifestyle; and it may even be a service to majority-language speakers to now and then encounter a message that they cannot understand—to realize that there are other conversations in which they cannot interpose.

And the "courtesy" that restricts language use to the ones that the majority know can be disastrous if the language suppressed is not spoken much anyway. This is a serious problem in some communities with endangered languages, where courtesy to outsiders means that children often do not hear the language spoken in public contexts. Implicitly, they draw their own conclusions about what the old language is good for. But that is another story.

It is not easy for every language to keep up its own community happenings in a Web environment. First of all, there needs to be a way for the written form of the language to be stored and transmitted. Even English can be a victim of the 7-bit straitjacket that the early ASCII pioneers imposed on language coding. How to coöperate with the naïveté of a belovèd hyæna, after all? There are very few European languages which can tolerate 7-bit ASCII without serious truncation. Probably only Dutch, Cornish, Basque, and Latin, although none of these except the last is a perfect fit. Going further afield, Indonesian, Swahili, Maori, Warlpiri (in Australia), Tok Pisin (in Papua-New Guinea), Nahuatl, and Quechua may join the minimalist band, although this is discounting the tendency of the languages of the Latin American zone to bring in tildes (ñ) and Spanish punctuation (¡¿). But there seem to be a fair number more in sub-Saharan Africa, the Pacific islands, and Australia. The out-of-the-way nature of such languages (and the fact that they exclude all the major European languages, as well as most others with populations over a million) is a little-noticed measure of how tyrannous a standard 7-bit ASCII really is. But it is a pleasant irony that it puts English on a par with a clutch of the little guys.

The "courtesy" that restricts language use to the ones that the majority know can be disastrous if the language suppressed is not spoken much anyway. This is a serious problem in some communities with endangered languages, where courtesy to outsiders means that children often do not hear the language spoken in public contexts.

Although Unicode has been defined, a basis at last to allow most languages' distinct characters to be specified independently of context, this is not yet in widespread use. But there are implemented standards now which respect the major character sets, at least when one specifies in advance which language's characters are intended. Even where they are not available, it is always possible to implement your own. In practice, it is easiest to implement this effectively when there is only a small community, and hence a single proposal can be propagated to all concerned. An example of this was the Naskapi people, who number fewer than a thousand, in the north of Quebec. In 1988-89, Bill Jancewicz developed a custom font and keyboarding system based on the Naskapi variant of Cree syllabary—this uses oriented geometrical shapes, and has been used for languages in the Great Lakes area since its invention in 1840. The Naskapi syllabary is now in use on over 20 PCs and 10 Macintosh computers, giving the Naskapi community a printed literature.

But even when you can mediate happily between characters on your keyboard and characters on your screen, there may still be problems in exchanging them with others. Strangely, as I discovered recently when attempting to institute correspondence in Russian, there is still no compatibility between Eudora mail systems and the universal Cyrillic coding system, KOI-8. This seems to be the manifestation of a general unpredictability of what is implemented and available when incorporating Cyrillic into application programs.

This kind of prestandardized existence is still typical of computerized versions of major non-European languages. It is being confronted head-on by the MILLE project at Lancaster University, a study of what is available in the way of language resources for the nonindigenous minority languages of the UK. MILLE focuses on Indian languages, such as Punjabi and Gujarati; but too often even when computerized applications are available to handle text, they do not use standard character coding, and so are not at all interoperable. The classic approach to providing support environments for text processing in new languages is:

1. a standard coding of its writing system;
2. localized versions of major operating systems and application programs;
3. the collection of a large text corpus;
4. the derivation of machine-readable dictionary and thesaurus materials.

(continued on page 44)

Fighting Words (continued from page 39)

After that, the language has the basics; and a number of other devices can be explored, such as parsers, multilingual dictionaries, and perhaps computer-aided language-learning materials. At New Mexico State University, a current research project is developing an architecture for working on what they call "low-density languages," i.e., ones where there is not much computer-readable corpus material from which to get started.

At Granada in May 1998, there was a workshop on Language Resources for European Minority Languages which made it possible to see the degree of progress for different languages along these now tested lines of development. From this it was clear that the major determinant of progress is the political autonomy of the community whose language is represented. Catalan and Galician lead the way, followed by Basque and Welsh, but Breton may be left behind. Even this may change in the near future; in October 1998, the French Government announced that it plans to sign and ratify the European Charter for Regional and Minority Languages.

All over the world, little by little, the work is being done. The foundations are being created for a much more multilingual computing future, where no language need be excluded—and hence, a more peaceful future for all of us.

Nicholas Ostler works for Linguacubun Ltd., and is the president of the Foundation for Endangered Languages. Email him at nostler@chibcha.demon.co.uk