

For Charles Ess, ed.

Cultural Attitudes Towards Technology and Communication

(New York: Suny Press, 1999)

Language, Power, and Software

Kenneth Keniston

Massachusetts Institute of Technology

In discussions of the impact of "The Information Age", the role of language in computing is rarely mentioned. Hundreds of books have analyzed the digital age, the networked society, the cyberworld, computer-mediated-communications (CMC), the impact of the new electronic media with hardly a word about the central importance of language in the Information Age.

The goal of this paper is to give language - by which I mean the language in which computing is done and in which computer-mediated-communication occurs - a key place in discussions of the impact of computation and computer-mediated-communications. I will argue that the language in which computing takes place is a critical variable in determining who benefits, who loses, who gains, who is excluded, who is included - in short, how the Information Age impacts the peoples and the cultures of the world. In other words, I will stress the relationship of language to power, wealth, privilege, and access to desired resources.

Localization and Language.

Although the ultimate "language" of the computer consists of digital zeroes and ones, the language of users, including programmers, is and must be one of the thousands of existing

languages of the world. In fact, however, virtually all programming languages, all operating systems, and most applications are written originally in English, making language a "non-issue" for the approximately seven percent of the world's population that speaks, reads and writes fluent English.

Since all major operating systems and applications are written in English (with the exception of the systems written for the German firm, SAP, which specializes in accounting software), use by non-English speakers requires localization. Localization entails adapting software written in one language for members of one culture to another language for members of another culture. It is sometimes thought to be simply a matter of translation. But in fact, it involves not only translation of individual words, but deeper modifications of computer codes involving scrolling patterns, character sets, box sizes, dates, dictionary search patterns, icons, et cetera. Arabic and Hebrew scroll from right to left, unlike the North European languages. Russian, Greek, Persian and Hindi involve non-Roman character sets. Ideographic, non-phonetic written languages like Chinese and Japanese involve tens of thousands of distinct characters.

Translation alone is an exceedingly complex part of localization. Ideally, it is a multistage process involving initial translation, followed by "back-translation" into the original language, comparison of the back-translated text with the original, adjustment of the translation as necessary, and incorporation of the now corrected translation into the final localized program. The cost per word thus translated has been estimated as approximately one dollar. Given that large programs like operating systems or office suites may contain tens of thousands of pages of text, localization even at the level of translation is both complex and expensive.

But localization involves more than simple translation. Scrolling patterns, character sets, box sizes, dates and icons must be adapted to the new language and the culture in which it is spoken. As one observer has noted with regard to computer icons, there is no gesture of the human hand which is not obscene in some language. As others have noted, the color red, which indicates "stop" or "danger" in the U.S., may indicate life or hope in another culture. Dictionary search patterns in a language like Finnish, which is highly inflected, require searching out the root verb from a word which may contain as prefixes and suffixes what in English would be the balance of an entire complex sentence.²

Moreover, localization is a worldwide business of growing economic importance. The industry association, the Localization Industry Standards Association (LISA) in Geneva holds periodic meetings of localizers and publishes a newsletter.³ Every major software firm has a localization division, and many attribute large parts of their sales not to the original English language version, but to localized versions sold in other countries. More than half of Microsoft sales are outside the United States - although not necessarily in languages other than English. As an industry, the localization industry is highly diverse and not geographically concentrated.

Other than the localization divisions of major software firms, there are literally hundreds of firms, scattered throughout the world depending on the linguistic area, which "specialize" in localization, often on subcontract from major software producers. Indeed the software giants of the U.S. often turn to small partners abroad to localize, or to test localized versions of, their

major packages. To my knowledge there is no study of the history and organization of the localization industry.

Localization is ordinarily seen as primarily a technical task. The localizer must not only be an experienced code writer, but must have a thorough knowledge of two languages, and ideally, of two cultures. Even localization from one North European language to another (e.g., from English to Spanish) requires good coding ability together with a knowledge of the subtleties of both languages.

"Localization" is intimately linked to another issue, commonly termed "standardization of code." To understand the importance of standardization requires analyzing how computers interpret letters - the letters, say, of standard English. Since computers can deal only with digital numbers, American computer coders early decided that the letters of the English language (along with numbers, punctuation marks, et cetera) would be mapped onto an eight-bit grid (which contained 256 theoretical possibilities). The standard known as ASCII (American Standard Code for the Interpretation of Information) assigns to each letter, number, and punctuation mark a specific numbered place among the 256 possible places. Thus, for example, the letter "lower case a" might be assigned location number 27, "lower case b," 28, et cetera. Computers, which communicate only in binary numbers, indicate first that an alphanumeric symbol is contained in the eight-bit word, and the decoding software then "reads" from a positive sign in location 27 the letter 'a', which it displays as an 'a' on the screen, adds to another word, prints as an 'a', et cetera. Communication between two computers is possible when they all use the same standardized code, such as ASCII. ASCII emerged to solve the problem of lack of standardization. In an

earlier period, each software manufacturer devised his or her own proprietary system for alphanumeric coding. Thus, one system's 'a' may have been location 27, while another's was location 203. Cross-platform intelligibility was impossible; each proprietary system required mastery of its own internal code; communication between two computers using different codes was impossible (or required complex transliteration programs). To solve this problem of a Tower of Babel, ASCII was developed and little by little imposed by its success on virtually all American software writers, and then, with modifications, on other languages whose characters could be adapted to the eight bit ASCII system. With modifications, ASCII, or a comparable eight bit (one byte) system, has proved adaptable to most languages except the ideographic languages like Chinese, which require tens of thousands of characters. For them, two-byte codes are necessary, involving 2562 possibilities. The emerging standard called Unicode, which aims at including all human languages, is a two-byte system.

But localization - whether it occurs, how it occurs, and how well and deeply it is done - is also an area where technology meets politics and culture in ways that I will emphasize in this paper. Elsewhere⁴ I have pointed to the ways that implicitly embedded cultural assumptions of the original language (almost always English) may (even in well localized software) be perceived as alien, hostile, or unintelligible to users in another culture. Here I will focus on the prior question of whether or not localized software exists at all.

Localization, or more generally language, has rarely been treated as an important topic in the literature on the impacts of the so-called Computer Age. But both individuals and governments have been acutely aware of this problem. The Indian high school student in Delhi with a perfect

knowledge of Hindi but a less than perfect knowledge of English confronts the issue of localization daily when he struggles with the "help" menus of his Windows 98 operating system - in English. The government of the tiny island republic of Iceland (population 500,000) confronts the issue of localization directly when it pleads with Microsoft to develop an Icelandic version of Microsoft's operating systems on the grounds that in its absence, young Icelanders are losing fluency in their traditional language. Of all nations, France has been perhaps the most vigorous in insisting on localization. A former French foreign minister termed the effort to preserve the hegemony of French against English "a worldwide struggle," "which we, the French, are the first to appreciate." Allying themselves with French-speaking Canadians and French speakers in so-called "Francophonic Africa," the French have made systematic efforts to suppress the use of English and insist on French. Software imported to France and Web sites developed in that country must use French as a matter of law. For the French, the enemy is the "Anglophonic tide." These French concerns are shared, though often less articulately and less overtly, in other parts of the world. A senior German telecom official recently commented, off the record, that German concerns over the hegemony of English in the computer world were almost as intense as those of the French. "But," he added, "we let the French do the talking for us."

More important, worries about the "Anglophonic tide" in software merge with deeper worries about the power of so-called "Anglo Saxon culture" on local values. What is the impact on villagers in African hamlets when satellite television permits them to see "Dallas," even if dubbed in Hausa, Igbo, or Swahili? How do Indian villagers react to Indian MTV, brought to them via satellite courtesy of Star TV, and MC'd in English by a laid back young Indian with an

American accent? How does the spread of computers and computer-mediated-communication (Internet, Web) influence existing inequalities of power within each society? How does it influence the gap between the rich societies of the North and the poor societies of the South? And does the dominance of English as the language of computation, Internet and the World Wide Web contribute to undermining the vitality and richness of ancient, non-Anglo-Saxon, cultures, especially in Africa and Asia?

These questions are too rarely asked, perhaps because they have no simple answers. Yet if we agree that the new electronic technologies are the most innovative and powerful technologies of the new millenium, then these questions, however difficult, must be asked. How do the new electronic technologies affect existing inequalities within and between nations? How do they impact the cultural diversity of the world?

Information Technology in South Asia

The seven nations of South Asia are in some respects unique, in some respects important in themselves, and in some respects illustrative of problems faced by many other regions. The basic facts about South Asia are well known. Approximately 1/4 of the world's population (1.2 - 1.3 billion persons) lives in the seven nations of India, Pakistan, Bangladesh, Sri Lanka, Nepal, Bhutan, and the Maldives. An estimated 5% of this population speaks good English, giving the subcontinent the second largest English-speaking population in the world, ahead of Great Britain and led only by the United States. English language fiction today is strongly influenced, indeed perhaps dominated, by writers of South Asian origin.⁵ Indeed, the articulateness of educated

South Asians in English is legendary. For the English speaking segment of the South Asian population, computing, almost entirely founded on the English language presents no problems whatsoever, nor does computer-mediated-communication (email, Internet, Web) in English.

There are, however, approximately 1.2 billion people in the Asian subcontinent who do not speak (or more important from the point of view of computation, read and write) good English. To begin with, approximately half of the population of the subcontinent is not literate at all. Equally important, most of the vast literate population of the region is literate in some language and script other than English -- or for that matter other than French, German, Spanish, et cetera, languages for which localized software is available for all major operating systems and many important applications.

South Asia contains some of the world's largest linguistic groups: for example, Hindi with an estimated 400 million speakers (approximately the population of the European Union), Bengali with approximately 200 million, and languages like Telegu with 80 million (about equal to the population of Germany.)⁶ There are literally dozens of languages with more than a million speakers in South Asia. India alone recognizes 18 official languages. Most of these languages have a unique script, and most have important literary traditions, both oral and written, that go back millenia. Some languages are cognate: for example, Urdu and Hindi both derive from the Hindustani of the Northern Plains, the one Persianized and the other Sanskritized in accordance with the cultural and political dictates of their respective speakers and nations.

In India today, major linguistic conflicts are largely absent. The initial plan to impose Hindi as the national link language has been repeatedly abandoned in the face of resistance from non-Hindi-speaking Indians, especially in the Southern states. The Indian states have been organized along linguistic lines, while English is accepted as the lingua franca of the national legislature, the higher civil service, the higher (national) courts, most highly educated people, and most national and multi-national businesses.⁷ But in Pakistan linguistic issues were central in the split between East and West Pakistan (what is now Bangladesh); and conflict over the role of Urdu, Punjabi, Sindhi, and other languages continues in today's Pakistan. In Sri Lanka, the Sinhala- and the Tamil-speaking populations have deep and destructive conflicts. So any simple generality about the role of language in South Asia fails. In India language is largely a non-issue in the political sense; in other nations, it is a cause or symbol of violent political polarizations.

One fact is constant, however. Throughout the entire subcontinent, English is the language of wealth, privilege, and power. For this reason, in Karachi, Dacca, Delhi, Colombo, and Katmandhu, parents who can afford it commonly seek English-language instruction for their children, aspiring to fluency in English at least as a second language in order to open to their children access to positions of responsibility, wealth, privilege, and power in their own societies and abroad. An Indian colleague tells of Hindu-nationalist villages in the most fundamentalist areas of India where every fourth shop on the streets offers English language instruction.

That English is the language of power, wealth, prestige, and preferment in South Asia is no accident. As many have documented, in the 1830's the English policy-maker Macauley laid down the rules that guided English colonial educational work in India (and elsewhere) from the

start. His goal was to use the English language, and to import English pedagogic methods and content in order to create a leadership group of "brown skinned Englishman", infused with English cultural values and loyal to the Empire. For more than a century, in India as well as in English colonies in Africa, Singapore, Malaysia, Hong Kong, and elsewhere, this plan guided British colonial linguistic policy.

Lord Macauley was a complex figure, an imperialist to be sure, but one who foresaw the day when India would claim independence as what he termed the "proudest day" in Great Britain's history.⁸ Moreover, in his belief that learning a language meant acquiring a culture, he anticipated the thinking of many modern applied linguists. One need not believe that language is reality in order to acknowledge that each language makes it easy to say some things, difficult to say others, and impossible to say still others. In short, language shapes, organizes and structures what we can communicate, how we think and what we experience.⁹ I recently worked with an MIT student brought up in Korea who was losing his facility with the Korean language. I expressed my regret and urged him to keep up his fluency. He commented with perception, "It doesn't really matter, because I can still think Korean." In other words, he was asserting that knowing a language entails knowing a way of organizing reality.

If Macauley's policy succeeded linguistically at least with Indian elites, it failed dramatically in other ways. As the Independence movement of India and other former British colonies showed, that policy failed to imbue in the population of South Asia, and even in English-speaking elites, an undying love for British rule and Empire. Politically, Macauley's policy was a complete failure, even if culturally it was partially successful. Men like Gandhi and Nehru in India, or

Jinnah in Pakistan, attacked the British raj in exquisite English, which they had often learned in English public schools and universities. Indeed, some have even claimed that "Anglo-Saxon" values of fair play, equality, the rule of law and the dignity of all human beings paradoxically helped inspire the movements of Independence of the former British colonies.

Studies of the elites of South Asia are rare and incomplete. Clearly, these elites differ from nation to nation, from region to region, from city to city. The Urdu-speaking elite in Pakistan that resulted from Partition differs in important respects from the business elite of Bombay or the political elites of Delhi. Moreover, with dramatic changes underway in the subcontinent, generalizations valid a decade ago may be invalid today. Witness, for example, the rise of a new younger generation of entrepreneurs in India, fueled by the progressive "liberalization" of the economy. Witness, too, the emergence of an elite group of the "captains of the software industry," today India's largest source of export earnings.

But whatever the characteristics of elites in South Asian cities and nations, they tend to have one common characteristic. For membership in South Asian elites, English is not only useful, but it is virtually the only privileged route to power, the only reliable key to any reasonable hope of wealth, preferment and influence. In South Asia as in few other regions of the world, language and power are fused. To be sure, English plays a similar role in the distribution of wealth, power and influence in other former British colonies in Africa and Southeast Asia. Moreover, throughout the world, English is today the preferred language of commerce and science, a fact almost as true in North Europe as it is in South Asia. In South Asia, however, the fusion of language and power is almost total.

What makes this relevant for computation and the impact of the Information Age in South Asia, and what differentiates South Asia from many other parts of the world, is the near complete absence of localized software in any of the traditional languages of this vast and populous region. Efforts have been made to change this situation; many schemes for localizing programs, operating systems, and applications to vernacular languages exist; many creative people are working on this problem. But the fact remains that, as of early 1999, none of these "solutions" has achieved any widespread acceptance. There are more plans than achievements; the policies of the Indian Government vis-à-vis localization remain complex and confused. Despite multiple proclamations on the part of both public and private groups that they have achieved a solution to the localization problem, either these solutions do not work or they are not widely adopted.

The result is that South Asia - with its vast population, its enormous economic potential, its multiple ancient cultures and literatures, and the world's largest, rapidly growing middle-class - almost completely lacks readily available, affordable, usable vernacular software. To put it bluntly and perhaps to overstate the point, unless an Indian reads, speaks, and writes good English, she cannot use a computer, she cannot use email, she cannot access the Web. Despite the valiant efforts of many who have tried to change the situation, English is necessary.

Why Is There No Local Language Software?

Given that South Asia possesses almost a quarter of the world's population, we need to ask why there is no effective and diffused localized software. An answer requires examining different levels of the problem.

First is the question of why the efforts of software companies in this area have been so meager or so ineffective. At the governmental level, India has promoted two distinct groups concerned with local language software, the National Centre for Software Technology (NCST) in Mumbai, and the Centre for Development of Advanced Computing (CDAC) in Poona. Each has followed a different path toward localization, with CDAC the first to market. CDAC's solutions were initially based on hardware modifications (the so-called GIST card), and its word processing software was seen by some users as inadequate and antiquated. Furthermore, CDAC, although a government agency, initially sold its local-language software, warts and all, for prices that drove away potential purchasers of lesser means. NSCT, which currently works with Microsoft on developing Indian language fonts, has developed alternative means of coding Indian languages, which many viewed as more likely to prevail than those promoted by CDAC. In Delhi, many agencies were directly or indirectly involved with setting policies that affect Indian language computing, including a special Government of India agency to promote the use of Indian languages, the Department of Telecommunications, and the Regulatory Authority of India. Competition or non-communication between these groups often resulted in conflicting rules or incompatible standards. Early on, of course, Indian computer scientists fully recognized the need for standardization of the major Indian languages and developed a coding system termed ISCII. ISCII is currently seen as more or less adequate for the northern Indian languages (which are based on Sanskrit and of Indo-European origin), but it is criticized as inadequate for the southern

(Dravidian) languages. Indeed, a recent meeting of Tamil-speakers from India and other countries rejected the use of IISCI in favor of another, proprietary code.¹⁰

At the corporate level, too, efforts have also been ineffectual or non-existent. Microsoft, which controls 95% of the operating system business in India, has a number of collaborations like that with NCST, to develop Indian language capabilities for its programs. Microsoft has announced publicly that the next version of Windows NT (Windows 2000) will contain "locale coding" ability for two Indian languages, probably Hindi and Tamil. But "locale coding" is not localization. Rather, it involves the capacity to use a basic English language program such as Word in order to input and print another language. Thus, for example, locale coding for Hindi entails a system of keyboard mapping such that the individual can input Hindi characters (either phonetically or through direct (stick on) keys; an internal software architecture that recognizes, interprets, and organizes these characters for output, and a set of fonts for monitor display and printing utilizing Hindi (Devanagari) characters. Although it is a step in the direction of localization, locale coding for Hindi nonetheless requires the ability to operate Windows and Word in English, and, in the case of keyboard mapping that uses the Roman keyboard phonetically, knowledge of the Romanized phonetic versions of Hindi words. Although it permits English-speakers to use the computer as something like a Hindi typewriter, it presupposes an advanced level of English.

Other multinationals and Indian firms have taken steps in the direction of localization. The MacIntosh interface lends itself to localization, and Apple has been a pioneer in localizing to Indian languages. The pity is that MacIntoshes are virtually unheard of in India, where they have

less than one percent of the market. IBM announced in 1997 a Hindi version of MS-DOS. The pity here is that MS-DOS has not been used as a programming language or operating system for many years in most nations. Modular Technologies in Poona has a series of innovative products that permit the use of several Indian languages. BharatBhasa, organized by the brilliant computer scientist Harsh Kumar, has made available as freeware an overlay for Microsoft operating systems that permits their use in a number of Indian languages. The ironic pity here is that since BharatBhasa is freeware, distributors have no financial motivation to circulate it, and its use is still limited. Finally, with the advent of Internet, literally dozens of "Internet solutions" for Indian languages are available on the Web for free. The pity there, however, is that most of these solutions are mutually incompatible: if you have Hindi system A and I have Hindi system B, their coding of Hindi characters is different and we cannot communicate with each other.

In short, despite valiant and brilliant efforts to develop local language software, their impact has been restricted. Of the major players, only Microsoft and the Government of India have the clout to create universally shared standards for the Indian languages and to build the localized software that would use them. Microsoft has chosen to focus its efforts on distributing English language software to the potentially large English speaking Indian market, and, as noted, on developing locale coding for two or more Indian languages. The Government of India's efforts have been dispersed in a variety of activities, often brilliant but together not effective in creating widely-used local language software.

The fact thus remains that the Gujarati merchant who would like to computerize his operation so that he does not have to stay up until midnight balancing his books can find no small business

applications except in English. The grandson from Delhi studying in London who would like to send email to his Hindi-speaking grandmother in Delhi must do so in English or not at all. The dynamic major Indian software firms, oriented toward exports and services, have shown little interest in localization. The creative work done by many Indian individuals and groups has so far not produced effective applications in the major Indian languages. Even with regard to online Indian newspapers, most of which are not in English, the lack of standardization is consequential. Since few newspapers share the same coding of, for example, Hindi, for each Hindi newspaper on the Web, the Web user must download the separate set of proprietary fonts used by that newspaper.

Computers, Power, and Global Monoculture

In the spring of 1998, President William Clinton spoke at the Massachusetts Institute of Technology on the Information Age. He devoted the first part of his talk to the wonders and potentials of the new digital technology. He stressed how it opens doors, provides access to information, facilitates communication, aids commerce and education.

But in the second half of his talk, President Clinton pointed out that computers and computer-mediated-communication also have the potential to widen the gap between the computer "haves" and the computer "have nots." As the haves increase knowledge, power, and access to resources, the gap between them and those who are "computer-deprived" grows. In the United States, where at present almost half of all households have computers, and of them about half are connected to the Internet and the Web, those who benefit most from the Computer Age are those who already

possess the greatest resources, political power and wealth.¹¹ The "information-deprived" are those who are already deprived in many other ways as well. Clinton ended his address by suggesting that market forces alone would not be enough to remedy this gap: both public action and private commitment are required to make the benefits of the Computer Age accessible to all.

In countless respects, the situation in South Asia is different from that in the United States. But in one respect it is the same: in both parts of the world, access to computers is empowering, and inability to access computers perpetuates deprivation, exclusion, and poverty. Indeed, as a general maxim in the history of technology, new technologies are appropriated by those who have power, and deliberately or not, these technologies serve initially to extend the power of those who already have power. In this regard, electronic technologies simply follow an historic rule.

But in South Asia, this universal problem is compounded by the overlap of power and language. Members of Indian elites are almost invariably English-speaking; India's vast population of peasants, tribals, scheduled and backward castes - the excluded and deprived (many of them illiterate) - rarely know any but a few words of English. This convergence of language and power in India means that in special ways, the Information Age perpetuates the powers of the English-speaking elite; it widens the already large gap between those who now have both power and English, and the nineteen out of twenty Indians who have neither. No one planned it this way, but the dominance of English as a computer language helps perpetuate existing inequities in South Asia.

The second important issue stemming from the importance of English in computers in South Asia is the issue of cultural diversity versus an emerging global monoculture. The political scientist Benjamin Barber has recently argued that world culture is increasingly polarized around two extremes.¹² The first is what he calls "McWorld" - the cosmopolitan, international, consumerist, multinationalized, advertising-based culture of cable TV, popular magazines, Hollywood films -- a culture which aims at universal accessibility, in which billions watch the same World Cup finals, a culture where MTV (translated), dramatizations of the lives of imaginary American millionaires, CNN, and films like Titanic dominate and flatten local cultures, producing a thin but powerful layer of consumerist, advertiser-driven, entertainment-based, and perhaps in the last analysis, American-influenced culture with great popular (if lowest denominator) appeal, backed by enormous financial and technological resources. It almost goes without saying that this culture is, in origins and assumptions, predominantly English speaking. Its centers are the U.S., Britain, Australia, English-speaking Canada, and English speakers in nations and city-states like Hong Kong, Singapore, South Africa... and India.

In defining the power of this global monoculture, computers, Internet and the Web play a small but growing role. In South Asia, countless million Indians have access to cable television, while three or four million at most have computers, and of them, perhaps ten percent have access to Internet and the Web. The driving forces of Anglo-Saxon global monoculture are still television and film. But the dominance of English in computation is part of this broader picture, and its importance is likely to increase in the years ahead. With the liberalization of Internet service providers in India, with efforts to lower the costs of local telephone connections, with the plummeting price of computers, more and more Indians are likely to join the "wired" world.

Rates of Internet growth are higher in South Asia than in most English-speaking nations, although the starting base is low and there are virtually no non-English Web sites or Internet hosts in these nations. At the same time, however, the dominance of English as defining the wired world remains intractable: indeed, an article in Salon, the on-line Apple magazine, several years ago, spoke of "the English speaking Web."¹³ While some counter examples exist (Hongladarom, this volume), the world of computers and computer-mediated communication must be counted almost exclusively as McWorld, not of cultural local diversity.

The Japanese sociologist, Toru Nishigaki of the University of Tokyo, sees a global Anglo-Saxon monoculture ultimately based on the power of American entertainment and American values as threatening to marginalize all local cultures.¹⁴ He notes that a Japanese businessman who is fluent in Chinese and wishes to communicate with a Chinese partner must, today, first translate his thoughts into English, communicate them in English via Internet to his Chinese partner, who must in turn re-translate them into Chinese. Equally emblematic of the power of American culture is the power of American technology. Given the low cost and effectiveness of American communication technologies, it often proves less costly and more efficient to send a message from Bombay to Calcutta via satellite through the United States than directly across India.

At the opposite pole from McWorld, Barber sees the ugly side of fundamentalism, which he terms "Jihad." He persuasively claims that one reaction against the cosmopolitan, internationalist, multinational and consumer-driven culture of McWorld is a return to the allegedly fundamental truths and varieties of an ancient culture. War is justified as an emblem of identity, an expression of community, an end in itself. "Even when there is no shooting war,

there is fractiousness, secession, and the quest for ever smaller communities."15 At worst, this return is exclusionary and even, as in the case of Jihad, may require holy wars against the impure. Jihad imagines a world of cultural and/or ethnic purity from which foreign, cosmopolitan, and alien influences have been eliminated, and in which an imagined ancient culture thrives, isolated from the rest of the corrupt and corrupting world. It is the world of "ethnic cleansing."

What Barber discusses as Jihad, however, also in his view has a different and friendlier face, namely that of cultural diversity. And in no part of the world is cultural diversity more manifest than in South Asia, and especially in India. Communal, religious, and ethnic tensions indeed exist and led, at the moment of Independence, to the tragedies of Partition and to repeated episodes of communal violence. Yet the fact is that India is the second largest Islamic nation in the world, with more than 170 million Muslims living - 99.99% of the time - in relative harmony with their Hindu neighbors. India is also the most multilingual and multicultural major nation on earth. Linguistic and cultural divides have torn apart or threatened to dismember nations like the former Soviet Union, Yugoslavia, Czechoslovakia and Canada, but in India they have by and large been managed harmoniously. No subcontinent in the world possesses so rich and diverse a set of cultures as South Asia.

The preservation of cultural diversity in the world, and in South Asia in particular, is a high value, perhaps on a par with the reduction of inequity and the promotion of political freedom. Cultural diversity can, of course, be perverted into reactionary fundamentalism. But this is most likely when local cultures are deprecated, spurned, marginalized, viewed as inadequate, and

when their members experience exclusion, condescension, or discrimination because of their membership in the culture. There is, then, every reason to value local cultures and to seek to make information technology a medium for their preservation and enhancement, not an instrument in their marginalization.

Given strong arguments that would support the creation of robust local language software in the major languages of South Asia, we need to ask why so relatively little has been done, despite the many voices raised to encourage vernacular computing. After all, the World Bank estimates that in the year 2020, India will have the world's fourth largest economy and the world's largest population. It is, of course, a poor nation at present, but it is also a thriving democracy, a nation with 500 million literate men and women, a nation with a rapidly growing middle class, and a nation which is, as Bill Gates put it, a "rising software superpower." India has twice as many university graduates as the People's Republic of China, although much higher illiteracy rates. In short, India, and South Asia more generally, is a region where one could anticipate a rapidly growing market for local language software in the decades ahead. Yet as I have noted, few are responding to this emerging market. Instead, what appears to be a "Tower of Cyber Babel" may be emerging with regard to Internet communication, and vernacular software remains, at best, a niche market.

Why So Little Local Language Software?

Among the reasons for the relative absence of local language software, economic factors surely play a key role. Indeed, it is often said that were there a market, localized software would simply

appear. Indians as a group are poor; telephone penetration is low (and therefore Internet penetration is necessarily low).¹⁶ It can be argued that, given the fusion of language, wealth and power in India, there is simply no market (and perhaps no need) for software in any language other than English. Asked about localization to Indian languages, international software firms sometimes reply, "But everyone speaks English in India," by which of course they mean that the present market consists of people who speak English. If this is accepted, then to produce a localized version of a major operating system or office suite in Hindi would not only be extraordinarily expensive but useless, since "all computer users speak good English." The same is even more true for other South Asian languages, because each of them has fewer mother tongue speakers than Hindi.

A related economic factor is the prevailing export orientation of the Indian software industry. To be sure, both the software and hardware associations of India have put localization at the top of their list of priorities. They insist that the great expansion of computer use and Internet to come in India will be domestic. If it is domestic, of course localization is required. But in fact, the orientation of the highly successful Indian software firms has been, so far, service-based, export-oriented, and therefore English-language based. One of India's greatest assets, reproduced in no other developing country, is its vast number of highly educated English speaking computer designers and programmers. For this reason alone, nations like China, Russia, and Brazil, whatever their other strengths, will continue to find it difficult to compete with India in the software field.

These economic factors are powerful and in the short run decisive. But I am reminded of the story told by Harsh Kumar, the inventor of the localization system known as BharatBhasha. He tells of the two shoe salesmen who go to a remote Indian village with a population of 1,000 people. The first salesman returns to his home office depressed and discouraged. "It is hopeless," he says, "there isn't a single person who wears shoes in the entire village." The second, however, returns jubilant and optimistic. "A wonderful opportunity," he says, "we can sell a thousand pairs of shoes."

Kumar insists that in the case of vernacular software, the absence of demand is created partly by the absence of supply. To take his favorite example, there are in Bombay hundreds, indeed thousands, of Marathi- and Gujarati -speaking merchants who own two or three shops and who currently spend every night until midnight balancing their books. They have the means and the need for computers which could do the job for them and get them home three hours earlier. But they do not have the command of English necessary to use any of the existing English-language small business packages. Computer consultants to whom they might turn can only offer English-language solutions, which are useless for the Marathi- or Gujarati-speaking merchant. The absence of supply automatically means the absence of demand.

At the very least, then, we need to examine critically the argument that economic factors alone suffice to explain the absence of local language software. Indeed there is a self-confirming quality to many economic arguments. If one asks why, in 19th century Europe, there was no demand for video cassette recorders, the answer is simple: there were no video cassette recorders available. An analogous reply might go part way toward explaining the absence of demand for

local language software: there can be no demand for a product which does not exist, or whose existence and utility is unknown. If local language software is not developed, or invisible, then the international software companies that claim that "there is no demand" will inevitably be correct.

A second factor that stands in the way of local language software is the very complexity - cultural, political, bureaucratic - of South Asia. One leader of a major American software firm, asked about localization to Indian languages said, "Okay, but which languages?". This is a reasonable question, but it has an answer: "Start with Hindi, go on to Bengali, Urdu, Tamil, Marathi, Telegu, et cetera." All of these languages are spoken by populations orders of magnitude larger than the populations of many nations for which locale coding or localization is currently available, for example, Norway, Denmark, or Latvia. Forward-looking companies, anticipating the steady growth of the vast Indian market, would be well advised to anticipate this market by localizing to major Indian languages. The winners in the next ten or twenty years in the Indian domestic market will be the firms that provide access to computers, Internet, and the Web in local languages.

Yet the complexity of the linguistic scene in South Asia points to the problem non-Indians (and some Indians as well) have in dealing with the subcontinent. India contrasts in this regard with the relative simplicity at the level of politics and written language of the other great Asian power, the People's Republic of China. In the latter, it is possible for American software firms to make binding agreements in Beijing for the use of the standardized written language that is employed by 1.3 billion Chinese. In India, for the many reasons suggested above, this is utterly impossible.

Other factors contribute to the slowness with which Indians and non-Indians alike have responded to the apparent potential of local language software. Among these is the fusion of language and power that has been at the center of this paper. The powerful in India, Pakistan, Sri Lanka, and Bangladesh are almost invariably those whose command of English is most perfect. Not only have they no personal incentive to encourage local language software, but, on the contrary, insofar as there is a class (or caste) interest in retaining power, it will be undermined by facilitating computer access to the non-English speaking, less powerful (and in India lower-caste) groups that already threaten the political hegemony of traditional Indian elites. I do not mean to suggest a conscious conspiracy, but only to propose that providing local language software to outcasts, tribals, scheduled castes, backward groups, slum-dwellers and other non-English-speaking local groups is unlikely to be paramount among the priorities of the powerful English speaking elites in South Asia.

Two other non-economic factors were once suggested by the head of a dynamic Indian software firm,¹⁷ who commented critically on a talk I once gave on local language computing. "You left out two of the central factors," he said, "the role of the Brahminical tradition and our ambivalent love affair with the English." By the first, he meant the traditional Brahmin emphasis on spirituality, transcendence, and higher orders of thought and action, contrasted with a distaste for all that is polluted, earthly, and material. "We are happy doing mathematics, astronomy, philosophy, and computers," he said, "but writing programs in Telegu or Hindi for the masses seems to many a less noble activity than programming in English or collaborating with a top-notch multinational firm in Germany." As for the "ambivalent love affair with the English," he

referred to the embeddedness in modern Indian culture of formerly English games like cricket, the preservation amongst the Indian upper classes of clubs, schools, firms, institutions, and forms of government associated with the British, and above all, the continuing use of English as the prestige language of India. "It is one thing to program in English, which connects us to the wealthy, powerful and rich nations - to the rest of the world. But to program in Telegu, Tamil, or Marathi is to descend to the level of the street, to renounce the efforts of a century and a half to become English, to ally ourselves with the forces of primitivism in our nation."

I cannot judge the validity of these arguments, but their claim is clearly that in addition to economic calculations, cultural factors play a role in the absence of vernacular software.

What Is To Be Done?

If local language software is important, and if it is largely absent in South Asia, the obvious question is, What is to be done?

Many wise men and women in India and elsewhere have answers to this question; mine will be a summary of theirs. First, however, I must note my disqualification: the solution to the problem of local software will obviously not come from American academics, but from the collaboration of South Asians in both public and private sectors interested in this problem, and perhaps from alliances with the multinational firms that today dominate the software market in South Asia. Here I can only offer a few suggestions.

1. The long-term potentials of the South Asian market need to be more accurately assessed.

Although the present installed base of both telephones and computers is low in South Asia, the growth of the South Asian middle classes is rapid. Firms that project five, ten, or twenty years ahead are likely to be winners. Long-term projections could be the basis for rational economic investments in local language software.

2. In India, the role of the states will be central to localization. Existing policy in India requires the use of local languages in each state. As these states move toward the computerization of basic operations like electoral rolls, drivers' licenses, land records, or the interconnection by Internet of district offices, local language software will be necessary. This demand will probably precede and exceed the demand from individual computer owners. (In the United States, two-thirds of all PC sales are to institutions, not individuals.) Serving this market from the state governments will require major investments in local language software.

3. Standardization of language codes is a prerequisite for local language operating systems and applications. The Government of India, multinationals, and major Indian software firms need to cooperate in developing broadly accepted standards for the major Indian languages and in persuading programmers in India and abroad to use these same coding standards for each Indian language. ISCII may be adequate. But if, as some claim, ISCII has inadequacies, especially for the Southern Indian languages, then corrections need to be made rapidly. The standardization of local language codes needs to be a priority for the Government of India; and the several authorities of that Government that today deal with local language software need to be brought together and instructed to produce unified standards on a firm deadline.

4. Local language software and multimedia should be actively promoted both by the central Government of India and the governments of the local states. If local language "content" on Internet and the Web continues to be absent, this will be an insuperable obstacle to local language information exchange. One positive role of government is to encourage (and finance, through start up grants) projects that use local languages in education, in the development of databases, in Internet communication, and in multimedia Web-based projects. The current initiatives of the Government of Andhra Pradesh and Tamil Nadu stand as models of what other States and the Government of India might achieve.

Summary

The growing importance of digital technologies in South Asia reveals problems and opportunities for that region and lessons for other nations in the world. In South Asia are visible two issues critical for every nation on earth: how can the new electronic technologies be used to close, rather than widen, the gap between the powerful and the powerless, the privileged and the underprivileged? How can the new technologies be used to deepen, intensify and enrich the cultural diversity of the world rather than flatten or eliminate it? These questions come together with particular intensity in South Asia because of the fusion of power and language on that subcontinent. But by the same token, solutions that develop in South Asia will be relevant to the rest of the world. Just as India has been an example of how a developing nation can preserve democracy and cultural diversity, so South Asian solutions to the challenges of the Information Age could be a model for the rest of the world.

(May, 1999)

ENDNOTES

1. An earlier version of this paper was prepared for the Conference on Localization at the Center for Development of Advanced Computing, Poona, Maharashtra, India, in September, 1998. The research on which the paper is based is partially funded by a grant from the Nippon Electric Company, administered through the Provost's Fund at MIT. I am especially grateful to Patrick Hall of the Open University in England for his comments on this draft.

2. There is an extensive technical literature on localization. Typical are the works of Nadine Kano, *Developing International Software for Windows 95 and Windows NT* (Redmond, Washington: Microsoft Press, 1995), and P.A.V. Hall and R. Hudson, *Software Without Frontiers* (Chichester, England: John Wiley and Sons, 1997). A work that stresses cultural factors more than most is Elisa M. del Galdo and Jakob Nielson, *International User Interfaces* (New York: Wiley, 1996).

3. <http://www.LISA.unige.ch>

4. Kenneth Keniston, *Software Localization: Notes On Technology And Culture* (Working Paper #26, Program in Science, Technology, and Society, Massachusetts Institute of Technology, 1997.)

5. See Salman Rushdie and Elizabeth West (eds.), *Mirror Work: 50 Years of Indian Writing, 1947-1997* (New York: Henry Holt, 1999). For a contrary view that stresses the importance of fiction in Indian languages, see Pankaj Mishra, "A Spirit of Their Own," *New York Review of Books* (May 20, 1999) 47-53.

6. Data on the precise numbers of speakers of Indian languages, or for that matter of any other language, are complicated by several factors. One problem is the absence of agreement as to what is required for it to be said that an individual "speaks X language." How much fluency? How much ability to read and write? are required. Linguists offer no consistent answers to these questions. In a nation like India, where bi-, tri-, and quadrilingualism is common, the primary source for figures on Indian languages is Ethnologue. (See below.)

The second problem has to do with the inadequacy of studies of linguistic patterns and usage in South Asia. For example, the most comprehensive sources on linguistic patterns in South Asia are found in <http://www.sil.org/ethnologue/countries/India.html>. But this document often relies on out-of-date figures (e.g., 1961 figures for English in India). Using more current figures, it indicates an extraordinarily low figure of 180 million primary mother-tongue speakers of Hindi (1991) and 346 million total Hindi speakers including second language users (1994). The *World Almanac, 1999* (Mahwah, NJ: World Almanac Books, 1999) lists 366 million native and 486 million total Hindi speakers, compared with 341 million native and 508 total English speakers. (By this reckoning, Hindi has the second largest number of native speakers in the world, while

English has the second largest number of total speakers.) All are agreed that Mandarin Chinese is far and away the most widely used language.

Furthermore, with regard to languages like Mandarin or Hindi, no agreement exists on how to categorize dialects that may be mutually unintelligible variants of the "same" language or nominally different languages that are naturally intelligible. In India, some dialects of Hindi are said not to be mutually intelligible. And in South Asia, Hindi and Urdu derive from a common origin in spoken Hindustani. Urdu uses Persian script and has been deliberately "Persianized" by Muslims, and especially by Pakistani authorities, who have made Urdu a national language.

(Before Partition virtually no one within the present boundaries of Pakistan spoke Urdu.) Hindi, in contrast, uses Devanagari script and has been to varying degrees "Sanskritized." Jawaharlal Nehru, whose native tongue was Hindi, complained that he could neither read the Indian Constitution in Hindi nor understand the Hindi broadcasts on Radio India because of the excessive Sanskritization of that language. See Stanley Wolpert, *Nehru: A Tryst with Destiny* (New York: Oxford University Press, 1996). The continuing congruity between Urdu and Hindi is shown by the enjoyment of Urdu television by Hindi speakers in northern India, and vice versa, and even more tellingly by the February, 1998 visit of Prime Minister Vajpajee of India to Pakistan. He addressed an Urdu-speaking Pakistani audience in Hindi, and, according to reports, was perfectly understood by the audience because of continuing similarities between Hindi and Urdu.

Similar imprecision exists with regard to the percentage of Indians who "speak English." The figure of 5% (approximately 50 million) is commonly accepted. But one commentator recently

argued that only 2% "really" speak good English, while others have claimed that the percentage is as high as 10%. And for the purposes of computation, no one (to my knowledge) has studied how much proficiency in English is required in order to use a computer whose operating system, instructions, and interface are in English. Once again, some claim that one or two years of language training are adequate; other argue that in order to use any complex computer program, very high levels of English proficiency are needed. Finally, there is the question of English language email and English language content on the Web.

Despite all these uncertainties, the overall linguistic pattern in South Asia is clear. In India alone, 18 languages (including English and Sanskrit) are officially recognized. There are, according to the Ethnologue figures, 30 distinct languages in India with more than a million speakers. Certain linguistic groups like Hindi speakers are as large as the entire population of the European Union; Bengali, with an estimated 200 million speakers, is approximately as common as French, Italian, and German combined. There are probably more Telegu speakers in Andhra Pradesh than there are German speakers in the world. The linguistic diversity of India, Pakistan, Bangladesh, Nepal, Sri Lanka, and the other South Asian nations is thus extraordinary.

But it is not unprecedented: among industrialized countries, Canada, Belgium, and Spain, to say nothing of the former Yugoslavia and the former Soviet Union, have very large linguistic subcommunities. The great majority of sub-Saharan African nations like Nigeria, Kenya, or South Africa, have multiple linguistic communities. Indeed, the monolingual pattern of the United States, where more than 95% of all inhabitants speak good English, is highly exceptional and perhaps even unique on the world scale.

7. The linguistic history of South Asia is complex and largely unanalyzed. Early works by Joshua Fishman, Charles Ferguson, Jyotirindra Das Gupta, *Language Problems of Developing Nations* (New York: Wiley, 1968), by Jyotirindra Das Gupta, *Language Conflict and National Development: Group Politics and National Policy in India* (Berkeley: University of California Press, 1970) and by Paul Brass, *Language, Religion, and Politics in North India* (New York: Cambridge University Press, 1974) lay out general issues as of 25 years ago. David Laitin's *Language Repertoires and State Construction in Africa* (New York: Cambridge University Press 1992) focuses on Africa, but uses the Indian model of a colonial language, a national language, plus a local language as the paradigm for Africa as well. Laitin assumes that the colonial language (e.g., English or French) is part of the national linguistic repertoire, but in the case of India and presumably most African nations, this is true only of a small cosmopolitan elite.

More recent works include Robert D. King, *Nehru and the Language Politics of India* (Delhi: Oxford University Press, 1997), and Tariq Rahman's excellent work, *Language and Politics in Pakistan* (Karachi: Oxford University Press, 1996).

8. On the history of British English language policy in India, see Anthony Read and David Fisher, *The Proudest Day. India's Long Road to Independence* (New York: Norton, 1998) and Gauri Viswanathan, *Masks of Conquest. Literary Study and British Rule in India* (New York: Columbia University Press, 1989).

9. This is known to communication theorists as the Sapir-Whorf Hypothesis. See, for example, William B. Gudykunst and Young Yun Kim, *Communicating with Strangers: An Approach to Intercultural Communication* (New York: McGraw Hill, 1997) or E.M. Griffin, *A First Look at Communication Theory* (New York: McGraw Hill, 1994). I am indebted to Charles Ess for these references.

10. See <http://www.elcot.com/tamilnet99.htm> (International Seminar on the Use of Tamil in IT, Chennai, February 7-8, 1999).

11. Donald A. Schon, Bish Sanyal, and William J. Mitchell (eds.), *High Technology in Low Income Communities. Prospects for the Positive Use of Advanced Information Technology* (Cambridge, MA: MIT Press, 1999) and David Eisner, "The Social Impact of the Internet: A Commercial Perspective," draft presented at June 4, 1999 meeting of the National Academy of Science/Max Planck Institute Task Force on "Global Networks and Local Values." Eisner notes, "Seventy-five percent of households with incomes over \$75,000 own computers, yet only 10% of the poorest families in this country [the United States] have computers." Eisner cites data from USA Today (no date).

12. Benjamin R. Barber, *Jihad versus McWorld* (New York: Times Books, 1995).

13. David Brake, "The U.S. Wide Web," *Salon* (Apple on-line), Issue 30 (September 3-6, 1996).

14. See <http://lpe.iss.u-tokyo.ac.jp/>

15. Benjamin R. Barber, *Atlantic Monthly* (March, 1992), 60.

16. Ashok Jhunjhunwala, Bhaskar Ramamurthi, Timothy A. Gonsalves, "The Role of Technology in Telecom Expansion in India," *IEEE Communications Magazine*, (November, 1998).

17. Nayaran Murthy, CEO of Infosys, remarks made at the National Institute of Advanced Studies, Bangalore, India, March, 1997.