Japanese Orthography in the Computer Age

J. Marshall Unger
University of Hawaii

Modern Japanese writing makes use of a large inventory of Chinese characters. Computers that can output such characters and represent them internally may cost more than computers that need only handle standard alphanumeric data, but there is no difficulty designing and manufacturing them. On the other hand, although many ingenious systems for input have been devised, none comes close to matching the efficiency of touch typing. Why is this, and is there hope for a breakthrough in the years ahead? This paper attempts to answer both these questions by examining the linguistic and technical factors which are responsible for the intractability of the input problem, and the social factors which lend it a sense of urgency. Special care is taken to separate those aspects of the problem which are genuinely cultural in nature from those which are not. The conclusion is not encouraging: an efficient general-purpose input system is unlikely. The Japanese might be better off directing some of the capital they are now investing in the quest for artificial intelligence towards less glamorous pursuits such as fostering public acceptance of romanization for use in computer applications that demand fast, accurate, easily modified input. The political resolve for such an approach, however, seems to be lacking.

1. Why is Japanese Script a Problem?

Uenohara Michiyuki, senior vice president of Nippon Electric Company (NEC, one of Japan's leading computer manufacturers), once summarized the problem of handling Japanese-language data on computers as follows:

> People using European languages can engage in data-exchange and dialogue with machinery through typing with almost the speed of conversation. But this is impossible in the Japanese language. If oral input becomes possible, the [*sic*] handicap will be totally eliminated. Because of the phonetic simplicity of the Japanese language relative to European languages, oral input will give the Japanese an advantage, reversing the present situation. The world is made in such a way that advantages and disadvantages are always offsetting relationships.
>
> The ultimate mission of technology is to make up for disadvantages of human beings and society and bring them progress. If it ruins advantages that human beings have, it does not deserve to be called technology.
>
> A local culture is something that has been developed through the long history of the region and human life itself. It should not be altered because of technology. Rather technology must be altered to fit the local culture. If the

Visible Language, XVIII 3 (Summer 1984), pp. 238-253.
Author's address: Department of Japanese, University of Hawaii, Honolulu HA 96822.

> Japanese language is abolished for the sake of convenience of usage of computers, the Japanese will be deprived of their identity. (Gregory & Etori 1981:J40.)

There are four major claims in this statement — one valid and three invalid. Although not all Japanese would subscribe to them, they give a feel for the consensus of opinion in Japan regarding the orthography/computer problem.

The valid claim, implicit in the first paragraph, is that input is the only real problem of data processing using Japanese script. The output of Japanese text on printers or television-like hardware, and the internal manipulation of data representing the elements of Japanese orthography are merely engineering problems. Existing methods suffice, and future methods will be better, cheaper, or both. The hard part is devising a suitable code for the machine-internal representation of the orthographic elements, and encoding raw Japanese-language data so that machines can use it.

The three invalid claims are all explicit. The first is perhaps the most obvious. Nothing "is impossible in the Japanese language" because the Japanese language is not Japanese writing. Even if the *current writing system* were "abolished" (which hardly anyone is urging), the Japanese language would remain. The obvious technical problem of improving the speed at which one can input Japanese orthography must not be confused with the more fundamental problem of deciding, in each application, exactly what sort of data one needs to input and manipulate.[1]

Just as the Japanese language (as distinct from Japanese writing) is no barrier to using Japanese-language data on computers, so too is its alleged "phonetic simplicity" of no benefit. As already remarked, the chief difficulty in using Japanese data on computers is inputting it. When this involves transforming phonemic representations of Japanese utterances into traditional orthography, using a microphone rather than a keyset as the means of input not only does not make the process any easier, it actually slows it down by introducing extraneous (non-phonemic) information which must eventually be filtered from the signal.

Finally, to speak of technology as the mere handservant of culture is a distortion: the two are inextricably related.[2] Changes in an unrelated field can suddenly wipe out the usefulness of a seemingly indispensable tool (e.g., the integrated circuit killed the slide rule) or breathe life into an abandoned one (e.g., the Arab oil embargo resurrected the New England wood-burning stove). The alphabet is not merely a cultural property; in a larger view of human history, it is a technological breakthrough. The question is not whether Japanese script can be handled *at all* on computers—no one ever doubted that—but whether it can be handled in a

way comparable in cost and efficiency to what would suffice if alphabetic writing were used instead.

2. Linguistic Factors

So far I have said nothing about the specific features of the Japanese writing system which make it unwieldy on computers. A comprehensive exposition of the writing system would require an article all its own,[3] especially if it were to be of value to readers unacquainted with spoken as well as written Japanese. But it is possible to get a feeling for the complexities of Japanese script without knowing too much about its "fine structure" because almost all of them arise from a single cause: the use of kanji.

Kanji are the Chinese characters which are intermixed with syllabic letters called kana in modern Japanese writing. Kanji are completely characterized by their relationships with one another, their symbolic values, and their shapes. These properties are illustrated within the ellipse in Figure 1. The various disadvantages that result from using kanji rather than phonemic notation to represent Japanese-language data on computers are categorized with respect to these three properties in Figure 2. Keep in mind that we are not concerned here with applications in which kanji are treated as data in their own right; in such "print shop" situations, one is merely interested in creating and formatting displays–the semantic content of the displays is immaterial. In computer applications, on the other hand, one usually wants the machine-internal data to be, to some extent, intelligible as language. Thus, only the last three of the following nine problem areas are related to the typographic properties of kanji.

1. Practical kanji inventories must be large. Most of the problems discussed below have their roots in this fundamental requirement. The Japan Industrial Standard (JIS), for example, calls for two sets (high and low frequency) of roughly three-thousand kanji each. Newspapers and book publishers maintain fonts of thousands more since the JIS sets do not include many rarely used and variant kanji.

2. In principle, any Chinese character can be pressed into service in writing Japanese. Even when there is a "standard" inventory of kanji available, provision must usually be made for *ad hoc* additions. The lack of a true standard, based on linguistic principles, complicates software design.

3. Alphabetic order is, of course, arbitrary; but the alphabet is short, and the process of alphabetizing distinct words is strictly mechanical. There is no similar simple algorithm for ordering words written in kanji. Since sorting and retrieving data from sorted lists is a basic task in

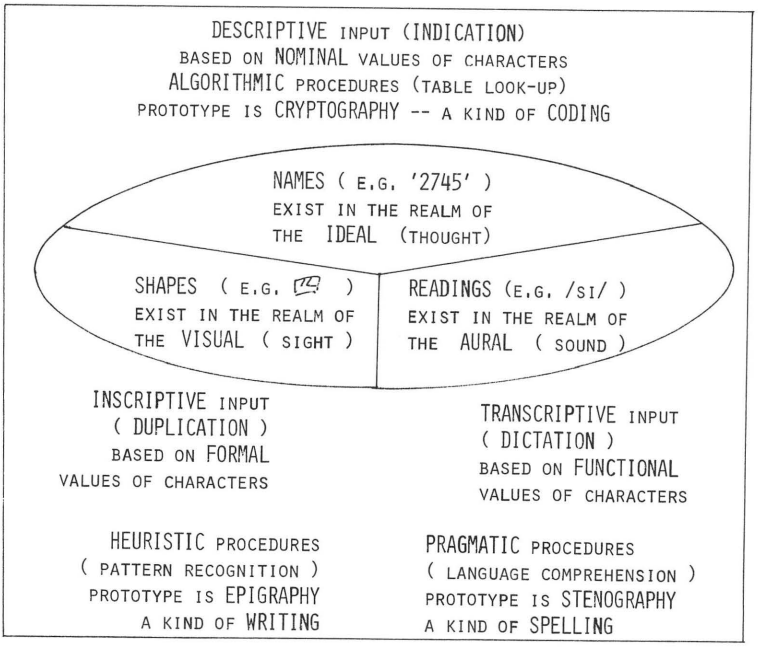

Figure 1.

Figure 2.

USING KANJI TO REPRESENT JAPANESE-LANGUAGE DATA ON COMPUTERS IS

BECAUSE	WASTEFUL	IMPRECISE	DIFFICULT
SETS OF KANJI ARE	(1) LARGE	(2) OPEN	(3) ILL-ORDERED
THE READINGS OF KANJI ARE	(4) REDUNDANT	(5) AMBIGUOUS	(6) ARTIFICIAL
KANJI SHAPES ARE	(7) COMPLEX	(8) ABSTRACT	(9) HOMOGENOUS

countless computer applications, the lack of a collating sequence poses an obvious difficulty.

4. The system of "readings" which relate kanji to morphemes of Japanese is highly inefficient because the vast majority of kanji occur very infrequently in ordinary texts.[4] Personal and place names contain hundreds which are otherwise not used at all.

5. Kanji which do occur frequently typically stand for two or more morphemes. Even in context it is not always possible to know with complete certainty the exact words represented by certain sequences of kanji. Proper nouns are notorious in this regard. Further ambiguities arise when kana are used after kanji to represent the endings of inflected words (verbs, adjectives, etc.). Although there are prescribed "spelling rules" for such cases, considerable variation is found in practice.

6. Some strings of syllables ("readings") are associated with only one kanji, but most are not. Dozens of kanji can, in the right context, substitute for the same phonemic strings. This has led to a plethora of Sino-Japanese homonyms, and the use of different kanji to distinguish different senses of the same word. In addition, many kanji usages are optional: kana may be used instead for stylistic reasons.

7. Kanji shapes are wasteful of system resources relative to standard alphanumerics because they require higher resolution for displays of the same quality and use up much more memory per character. Falling hardware costs are narrowing the gap between systems that can and cannot output kanji, but that gap will never be reduced to zero.

8. Many kanji have variant shapes; recent reforms (different in China and Japan) have added hundreds more. These are not matters of font style but variations in the number, size, relative location, and scale of the graphic elements which are used in combination to form the vast majority of kanji images (*jitai*).[5] Thus, graphically different kanji may have the same symbolic value. In order to identify allographs, the best one can do is try heuristic or brute-force methods of the kind used in chess-playing programs.

9. Each of the basic graphic elements that combine to form kanji images have considerable visual redundancy, but kanji often differ only in one of these elements. Certainly 'i' and 'j' look similar; but when one is dealing with thousands of symbols, optical scanning begins to get intolerably difficult as such "near misses" multiply. And, of course, handwriting makes matters much worse.

Although the nine characteristics of kanji just enumerated present difficulties for kanji use on computers, many resulted from or allowed for the successful use of kanji in Japan for centuries.[6] People have little

difficulty, for example, identifying traditional variant forms of kanji with their mental norms because, unlike machines, they perceive kanji as complete *Gestalten* situated in context. They easily accomodate themselves to high levels of ambiguity and artificiality in kanji/morpheme relationships far beyond what existing input programs can tolerate. They have devised ingenious tricks for sorting and retrieving Japanese data written in kanji because they take in *signe* and *signifié* holistically.[7] Thus, the kanji problem has nothing to do with the cognitive difficulties of learning and using kanji to write Japanese. Whatever the difficulties, people obviously cope with them. The problem is precisely that machines don't. Our tendency to associate literacy with culture makes us forget that a writing system is first and foremost a technology. Once this fact is recalled, it becomes clear that what we have in the case of Japanese data processing is a conflict between an ancient technique for reducing the Japanese language to writing and a modern technique for reducing writing to numbers.

It is important to stress this point for two reasons: first, because there are people (like Uenohara) who mistakenly think the conflict is cultural rather than technological; and second, because there *are* intercultural pressures against kanji use which must not be confused with the very specific problem of handling Japanese orthography on computers. For example, the older practice of making up Sino-Japanese compounds such as *jidōsha* 'automobile' and *denwa* 'telephone' has given way to deluge of Japanese-English that makes Franglais seem like a summer shower. Kabashima Tadao (1981:178ff.) has pointed out an important reason for this: part of the post-war orthography reform involved banning many *kun-yomi* or native glosses for kanji. Thus the percentage of kanji that had only *on-yomi* (i.e., could represent only Sino-Japanese morphemes) rose dramatically, and Japanese school-children were suddenly deprived of the mnemonics that had aided past generations of learners. For them a Sino-Japanese neologism is more often than not a troublesome piece of pedantry rather than a meaningful part of their inner symbolic landscape.

Indeed, kanji exist on a semiotic level congruent with neither the phonological nor the semantic structure of the Japanese language. When Japanese say that writing Japanese without kanji is inconceivable, they are referring, I believe, to this stratum of extra meaning. It is non-linguistic but it is not an illusion; one must not ignore its existence even if some Japanese (e.g., Suzuki 1977) wrongly insist that it is part of the synchronic structure of the Japanese language. In fact, the phenomenon of non-linguistic meaning that is experienced when one reads Chinese or Japanese is, I think, a hitherto overlooked piece of evidence against the Platonist/rationalist perspective now favored by many linguists. This is the

epistemological view which underlies rigorously formalized theories of grammar, much work on machine translation, and above all artificial intelligence research. Ironically, as we shall see below, the future of the kanji input problem is tied closely to just such applied research.

3. Technical Problems

As remarked at the beginning, the nine points summarized in Figure 2 are merely inconveniences to the extent they pertain to output and machine-internal calculation; but to the extent they relate to input and the problem of devising a suitable code for kanji, they are serious obstacles to progress.[8] Liu (1983:415) estimates that in China alone "nearly 400 schemes" for character encoding have been proposed, of which "several dozens" have actually been implemented. Many of these coding schemes may differ only slightly from one another, but distinctive systems developed in Japan, Taiwan, Hong Kong, the US, and other countries should also be taken into account; clearly, an enormous amount of work is going into input and encoding of characters.

Figure 1 gives a typological summary of Chinese-character input systems. The inclusion of kana in Japanese texts makes the input of Japanese different in some respects from the input of Chinese; but the methods available for both languages are essentially the same, and all are based on these three basic types in one way or another.

Inscriptive input, in which the user actually "draws" characters on a digitizer or similar device, is the least developed of the three because pattern recognition is such a hard computational problem, and because even a perfect system could not proceed faster than ordinary handwriting. Research in this area is now focused primarily on optical character recognition (OCR). Its goal is to produce machines that can scan printed documents and transfer them character-by-character to electronic media with minimum user intervention.

Transcriptive input is now the "method of choice" in Japan. The user types strings of kana or roman letters and instructs the computer to convert portions of them into kanji. Although some programs use frequency statistics or elementary syntactic analysis to narrow down the number of alternative kanji, the user must often choose among them. This means that, if a text is being copied, the user must momentarily look away from it. A maximum rate of about 50-60 characters per minute can be achieved this way, though there are claims for higher rates when ergonomically designed keysets are used in conjunction with this kind of input. As already mentioned, using voice input rather than typing for transcriptive input does nothing to ameliorate its inherent shortcomings.[9]

Descriptive input has more variants than the other two methods. There

are four subtypes depending on whether the codes for the kanji are mnemonic or arbitrary, and whether the hardware implementation uses a small or a large keyset. Mnemonic codes may make reference to kanji shapes and/or readings. Large keysets (both "tablets" and "multi-shift" keyboards) almost invariably use some sort of mnemonic arrangement. Characters may be grouped according to relative individual frequency, co-occurrence in pairs, shared graphic components (radicals), common readings, and so forth. On small keysets (standard or specially designed keyboards with roughly the same number of keys as a typewriter), the individual keys almost always stand for primitive elements (phonemes and syllables, strokes and radicals) into which kanji readings or shapes can be analyzed. The idea is that the user should be able to derive the code for any available kanji by applying a simple set of rules; this presumably makes the system easy to learn and efficient to use. In practice, however, the reported results for small-keyset systems are 120 characters per minute or less.

Transcriptive and many descriptive input methods are really two-stage processes. During stage one, the user inputs raw character codes. Stage two involves disambiguating kanji that share the same code. Novices generally interrupt stage-one work and attend to stage-two work whenever an ambiguity arises. More experienced users often try to type in a whole paragraph or page "blind" and go back to alter incorrect kanji. But the twofold nature of the process cannot be eliminated entirely; as long as the codes are mnemonic, they will be ambiguous to some extent. Stated differently, kanji will always cost more to process because they involve extra-linguistic information that cannot be automatically predicted from raw speech or an abbreviated form of writing.

The bifurcation of the input process, however, is not the only way to handle this extra cost. Another approach is to use completely arbitrary codes, thereby sacrificing "ease of learning" for the benefits that come with true touch typing.[10] In English beginning typists are expected to be able to handle 55 words per minute (five strokes = one word; each error = -1 word). Experienced typists often break 100 words per minute, and the Dvorak Simplified Keyboard puts such rates within the grasp of people of only average ability. Assuming two-stroke codes for each high-frequency character, with a smaller number of three-stroke codes for uncommon kanji, we would therefore look for a rate of two-hundred or more characters per minute for Japanese. As already stated, existing input techniques don't come close to this; moreover, those that require constant user intervention (stage-two work) cause excessive fatigue, muscular discomfort, and eyestrain. The extra labor necessitated by kanji cannot be eliminated; but undesirable side-effects can be avoided by shifting this

extra effort to the initial skill-acquisition phase of user activity and away from *ad hoc* disambiguation of characters.

Yamada Hisao and his colleagues at the University of Tokyo are well-aware of this trade-off. They are concerned with the problems faced by office-workers who input Japanese text five to six hours a day, five or more days a week. With such typists in mind, they have devised a keyboard arrangement in which the most frequent 2,304 (= 48 squared) kanji, kana, and punctuation marks are represented by two-stroke sequences assigned solely to optimize input speed (Hiraga et al. 1980). The keys have no names or meanings in this system; a map of the kanji and kana in relation to key pairs appears to be totally patternless. Intuitively, such a scheme should be hard to learn; but experimental results suggest quite the opposite. Yamada's hypothesis is, in fact, that no mnemonic coding scheme can be optimal. Any reference to the shape or function of kanji actually slows down input because it interferes with those subliminal kinesthetic skills that make touch typing viable. To use a popular metaphor, mnemonic codes access so-called "left-hemisphere" cognitive functions whereas touch typing depends on concentrated "right-hemisphere" processing.[11]

This brings us to the boundary between technology and economics. The hardware and software needed to implement Yamada's touch typing input system is all "old technology," well understood, widely available, and cheap. By contrast, the transcriptive kana-to-kanji and roman-to-kanji conversion systems being pushed by the major computer manufacturers in Japan right now require extra memory for look-up dictionaries, elaborate software, and sometimes non-standard keysets as well. Thus, if a touch typing standard is to be adopted, it will have to be adopted soon; otherwise manufacturers and their customers will have too much invested in research and develpment costs, outmoded machines, and user training to give up transcriptive input.[12] On the other hand, if a touch typing standard is not adopted, Japanese offices, laboratories, and factories that cling to kanji will, it seems, have to process information about half as fast as the rest of the industrialized world, and pay a premium for the privilege of doing so. Finally, there is the possibility that a touch typing standard will be established, but that it will be found impractical for use except by the office-workers for whom it was originally conceived.

Under the circumstances, the only way out would seem to be rescue by a *deus ex machina*. And that is literally what appears to be on the horizon: Japanese government and industry are now collaborating on a project aimed at producing the world's first genuine artificial intelligence (AI) computers by 1990. These "supercomputers" will allegedly be capable of creative problem-solving, natural-language translation, general pattern

recognition, and other seemingly miraculous feats. According to some experts (e.g., Feigenbaum & McCorduck 1983), it is only a matter of time before such machines are built; others have questioned whether such machines are possible even in principle (Dreyfus 1979) or, if possible, whether they are always socially desirable (Weizenbaum 1976). In view of this controversy—and at this stage, I think it fair to say that the doubters have the preponderance of hard evidence on their side—it is natural to ask: to what extent is the ultimate solution of the kanji input problem dependent on the outcome of the fifth generation project?

I had a chance recently to pose this question informally to several Japanese experts attending the 1983 International Conference on Text Processing with a Large Character Set in Tokyo. Responses were of three kinds: the majority felt existing input techniques are essentially adequate and merely need to be refined; they see the outcome of the fifth generation project as unrelated. Most of the rest held that AI machines would be needed for a truly general and elegant solution of the kanji problem. A few, however, went further; in particular, Motooka Tōru of the University of Tokyo, chief planner of the fifth generation project and (significantly) the conference chairman, opined that there would still be difficulties with input *even after* AI machines becomeavailable. This is probably the most prudent view for two reasons. First, the majority consisted mostly of engineers, many of whom were personally involved in the production of systems on the market. Second, a certain amount of "not invented here" thinking (abetted by the Japanese penchant for professional factionalism) was in evidence. For example, it was members of the majority who most strongly criticized the goal of voice input, one of the items high on the agenda of the fifth generation project.

4. Social Aspects

At any rate, if there is a lack of unanimity among experts on the technical connection between the fifth generation project and the kanji problem, the historical evidence is clear enough. A recent "think-tank" report summarizes the situation this way:

> There is no dispute among industry participants that Japan lacks the software capability needed to fuel its computer sales. . . .The reason for this lag stems from two interrelated sources: (1) the lack of emphasis on national projects that needed high-level software capabilites, and (2) the low esteem of software engineers/programmers as a profession. For example, 20 years ago the United States launched its National Aeronautics and Space Administration (NASA) space programs, which acted as a catalyst for developing software systems. Conversely, . . . Japan had no comparable government-sponsored program and did not establish software priorities until the late 1970's. In part a reflection of the national focus (or the lack of it) and in part a reflection of cultural values, the

> Japanese education system, as a result, did not prepare the talent it needed to develop software. Although Japan has some 50 scientists and engineers per 10,000 in the labor force, less than 6 percent of them become software engineers or programmers. In actual numbers, Japan has only 7 percent of the number of software professionals in the United States. (Ohmae 1983:70-71.)

It goes without saying that the fifth generation project is exactly the kind of national endeavor that Japan neglected in the past. And the "cultural values" that contributed to that neglect are, of course, tied to kanji-based literacy.

Japanese study the reading and writing of their native language through nine years of compulsory education and three years of high school; those going on to university continue to study kanji, though on a less systematic basis. Anything that calls into question the value of their long years of effort in acquiring conventional literacy makes them feel uneasy, and they do not want to spend more time learning new ways of writing Japanese, however simple or useful they may be. On the other hand, they dislike having to work in a computerized world in which the flood of English-language data is growing daily. Given these conflicting emotions, it is easy to see how the fifth generation *deus ex machina* can grow from a dry intellectual hypothesis into an object of faith.

Of course, the Japanese are not the first to become infatuated with the dream of a disembodied Platonic intelligence. Already in 1976 Weizenbaum identified an American "artificial intelligentsia" that showed all the symptoms of chronic orthodoxy. One of its leading exponents, Edward A. Feigenbaum of Stanford University, has wasted no time in championing Japan's drive towards AI. Unlike Vogel *1979 (Japan As Number One)* and Ouchi 1981 *(Theory Z)*, bestsellers which obviously inspired it, *The Fifth Generation* (Feigenbaum & McCorduck 1983) contains hardly any solid information about its namesake; rather, it is an extended polemic in support of the authors' personal views on where AI research should be going. The message is unabashedly sensationalist: the Japanese are going to build genuine AI machines while the complacent, skeptical West slumbers; when it awakes, it will be too late—the Japanese will be in control of the key to the economic power of the future. The following passage illustrates the smug confidence in AI theory that informs the argument:

> It's a distinct pleasure to report that while the Japanese have put a lot of planning and thought into their Fifth Generation project, they've spent no time at all in those arid little debates so beloved by Western intellectuals, debates centered on the question whether a machine really can be said to think. They regard our obsession with that topic the same way we regard their eating raw fish—an odd, puzzling, but harmless cultural quirk. Instead, their debates are about the best way to design an intelligent machine, truly a new generation, the engine that will produce the new wealth of nations. (1983:17.)

I can only hope that the authors' quaint ideas about *sashimi* put readers on guard: the absence of "those arid little debates" at which Feigenbaum and McCorduck sneer may be a cause for rejoicing in some circles; but a more plausible interpretation is that Japan is eager to rush in where even IBM fears to tread at least in part because of intellectual shallowness.

Feigenbaum and McCorduck are so wrapped up in their self-congratulatory envy of the "forty samurai" spearheading the fifth generation project that they overlook its linguistic and sociopolitical aspects entirely. They discuss "Japanese Computer Science Education—an Achilles Heel?" (1983:144-147) without suspecting that the inadequacies they perceive are not confined just to computer science. "Another myth," they say, "goes something like this: *The Japanese might be able to build terrific cars, stereos, cameras, and baseball mits, but computers are different. The Japanese can't produce software. But it's not the fault of their intelligence; it's a limitation of their language.*" (1983: 141; emphasis in the original.) The only people who think that way seem to be Japanese (like Uenohara) who habitually confuse writing in kanji with thinking in their own language. In "Some Views from the Companies" (1983:108-110) Feigenbaum and McCorduck document the doubt and hostility with which some of Japan's hi-tech corporations view the national commitment to the fifth generation project, but they totally misinterpret this important fact. Companies vie fiercely for bright college graduates in Japan; the competition for computer science majors is particularly intense. Since many managers believe (as we have seen) that consumers will accept the latest crop of machines and programs, they are loathe to loan out their "whiz kids" to a project that, at best, will produce a prototype by 1990.

One must be careful, however, not to go to the other extreme and ignore the cultural aspects of computerization. The widespread belief that kanji are indispensable for writing Japanese should not be ignored even through it is rarely more than the result of misunderstanding or prejudice. To someone accustomed to reading mixed kanji/kana script, sentences without kanji do have a phantom-like, insubstantial quality about them; they seem somehow naked without the clothing of extra-linguistic meaning which kanji provide. As remarked at the end of Section 2, this intuitive feeling that the linguistic meaning of a written Japanese sentence needs to be completed by information unrelated to actual speech is one of those everyday (and therefore easily overlooked) cognitive experiences found in all cultures that cast doubt on the reductionist epistemology on which AI is based; yet ironically it is the very thing that makes so many Japanese want computers that can handle kanji. Without kanji, most Japanese feel they have no way to show others that they are educated and possess the knowledge which entitles them to social acceptance. Indeed,

education, knowledge, and kanji are often just one big blur in the Japanese mind.[13] Unless and until there is a perception that the leaders of Japanese society are prepared to foster a transition to the non-use of kanji, it is doubtful that many Japanese will be willing to take the risks involved in beginning that transition individually.

5. Conclusion

Thus, in looking ahead to the proliferation of computer power throughout Japanese society, our attention should, I think, be focused on government ministries rather than corporate laboratories. The bottleneck of kanji input simply cannot be widened. The obvious move is to use standard alphanumerics in all applications that do not require Japanese script, including real-time communications, creative rogramming, and so forth. Since it is unlikely that, even in the twenty-first century, the bulk of one's life will be spent in front of a computer terminal, the establishment of romanization as a special-purpose alternative mode of writing would probably not mean the end of traditional script. Unfortunately, it seems that many of the key "players" in Japan today are unwilling to countenance any compromise whatsoever.

1. Suzuki Takao (1977) contends that, at least in the case of Japanese, orthography is an integral part of what we should properly call language. It is obvious, however, that even if we were to grant some special psycholinguistic status to Japanese orthography, it would be of no relevance to machines in which all information is reduced to sequences of binary digits (bits).

2. Burke (1978) presents dozens of historical cases that illustrate this point. For example: "The change in the weather which struck northern Europe like a sledgehammer in the twelfth and thirteenth centuries provided urgent need for more efficient heating. The chimney answered the need, and in doing so had the most profound effect on the economic and cultural life of the continent." (1978:291)

3. For example, the introductory material in Hadamitzky & Spahn (1981)—a handbook for students starting to learn modern Japanese reading—runs to nearly seventy pages.

4. Between July 15 and July 21, 1971, a sample of 4,252 sentences on the Kyōdō News Agency wire were studied for frequency. Of the 1,001,554 character tokens, 462,209 or 46.1% were kanji; however, these kanji accounted for 2,258 or 86.8% of the 2,601 character types tallied (see Hayashi 1982:206).

5. According to Hayashi (1982:216) kanji composed of simpler graphic elements (*kaii moji)* and *keisei moji* accounted for 1,624 or 87.8% of the 1,850 *tōyō kanji*; the 95 new kanji in the *jōyō kanji* list are almost all of these two types.

6. Of course, "success" must here be understood in a relative sense. Tōdō (1982:170-175) comments on illiteracy in pre-war Japan and argues that the post-war orthography reform was as important for Japan's recovery as the land reform and new constitution. Miller (1982:186) questions the reliability of official

literacy statistics, claiming that no country could have 94.7% literacy simply because of blindness and other impairments.

7. These are precisely the aspects of human intelligence which Dreyfus (1979) argues are not formalizable and therefore permanently beyond the grasp of digital computers.

8. I hasten to add, however, that even if input is left out of consideration, the extra costs of kanji output and manipulation are substantial relative to the alternative of using standard alphanumerics.

9. Confusion on this point is not confined to Japan. Panko (1983) is an example of an "in depth" article by an American who, though fairly knowledgeable about the linguistic facts of Japanese life, mistakenly seizes on the keyboard as the nub of the kanji input problem.

10. Yamada (1980) summarizes the development of touch typing in the West and its implications for Japan. Touch typing is much more than a mere convenience; it is a remarkable breakthrough in its own right. "Consider a typical secretarial task, the retyping of a document . . . . [I]t closely resembles a series of choice-reaction-time tasks, in which a subject is presented with a single visual stimulus from a set of two or more stimuli after being instructed to rapidly press a particular button for each of the possible stimuli. Under optimal conditions . . . the average latency, or delay between the presentation of the stimulus and the pressing of a button, is approximately 250 milliseconds. The paradox of typing is that a latency of 250 milliseconds yields a typing rate of 48 words per minute . . . . Yet speeds of twice that rate are fairly common." (Salthouse 1984:128.)

11. In recent lectures Yamada gave two pieces of evidence which suggest that touch typing, once mastered, is strictly "right-hemisphere" (non-linguistic). One is the observation that many experienced American secretaries can accurately copy-type manuscripts while carrying on conversations. The other is the fact that many of these same secretaries prefer to place the manuscript to their left; this contradicts the practice of many typing schools, but does fit the "right-hemisphere" interpretation. I can think of two similar pieces of circumstantial evidence. First, it is hard to sightread the lyrics of a song and its piano accompniament at the same time; however, once either the music or the words have been learnt, it is easy to concentrate on the other, singing and playing at the same time. Second, skillful magicians never perform the same trick twice for the same spectators unless they can vary the method. In one classic effect, a spectator makes some simple moves with half a deck of cards, imitating what he sees the performer doing with the other half. This makes it impossible for the spectator to see through the trick because "the mind cannot think of two things at once while executing a manual operation which involves the use of the eyes . . . ." (Hugard & Crimmins 1961:74-75.)

12. In the West the Dvorak Simplified Keyboard has been available for nearly fifty years, but similar factors have kept it from replacing the grossly inefficient "qwerty" arrangement in common use.

13. Miller (1982) discusses this and related Japanese sociolinguistic attitudes at length. He attributes them to a "sustaining myth," a kind of obsession with the Japanese language, that has emerged in the post-war period as a psychosocial compensation for the misery and humiliation of the defeat. Although I agree with many of Miller's observations in a general way, I cannot accept mass psychological trauma as a valid explanation. (Indeed, Miller's arguments are so extreme and his tone so scathing that one suspects the "myth" is just an extended metaphor designed to allow the author to attack Japanese linguists he personally dislikes

under the guise of establishing a serious theory.) In my opinion, the attitudes which Miller catalogues are the result of the abolition of elite higher schools by the Allied Occupation, and the accompanying changes in public secondary education. A full discussion of this hypothesis is beyond the scope of this paper, but I think it important to mention one aspect of it here. Research on the *level* of literacy among the *great majority* of Japanese prior to 1945 is almost entirely lacking (Richard Rubinger, personal communication). Although compulsory egalitarian elementary education was instituted early in the Meiji Period, other evidence (such as *Tōdō's* remarks referred to in note 6) suggests that a thorough command of kanji of the sort now required of all Japanese high-school students was not expected of most Japanese before 1945. If this is so, then the blurring of kanji, education, and knowledge is a relatively recent phenomenon, and would account for the kind of attitudes which Miller attributes to his "myth."

References

Burke, James. 1978. *Connections*. Boston, Toronto: Little,Brown & Co.

Dreyfus, Hubert L. 1979. *What Computers Can't Do: The Limits of Artificial Intelligence*. Revised edition. New York: Harper Colophon.

Feigenbaum, Edward A., and Pamela McCorduck. 1983. *The Fifth Generation: Artificial Intelligence and Japan's Computer Challenge to the World*. Reading, Massachusetts: Addison-Wesley.

Gregory, Gene Adrian, and Etori Akio. 1981. "Japanese Technology Today: The Electronic Revolution Continues." Advertising supplement. *Scientific American*, October. 97-144.

Hadamitzky, Wolfgang, and Mark Spahn. 1981. *Kanji & Kana: A Handbook and Dictionary of the Japanese Writing System*. Rutland, Vermont: Charles E. Tuttle & Co.

Hayashi Oki, et al., comps. 1982. *Zusetsu Nihongo: Gurafu de miru kotoba no sugata* [Japanese in diagrams: The form of language seen through graphs]. Kadokawa shōjiten 9. Kadokawa shoten.

Hiraga Yuzuru, Ono Yoshihiko, and Yamada Hisao. 1980. "An Assignment of Key-Codes for a Japanese Character Keyboard." University of Tokyo Department of Information Science Technical Report 80-12 (July). In *Proceedings of the 8th International Conference on Computational Linguistics (Tōkyō)*, 249-256.

Hugard, Jean, and John J. Crimmins, Jr., eds. 1961. *Encyclopedia of Card Tricks*. London: Faber & Faber.

Kabashima Tadao. 1981. *Nihongo wa dō kawaru ka* [How will the Japanese language change?]. Iwanami shinsho 145. Iwanami shoten.

Liu Yongquan. 1983. "Language Engineering in China." In *Proceedings of the ICTP '83*, 412-418.

Miller, Roy Andrew. 1983. *Japan's Modern Myth: The Language and Beyond*. New York & Tōkyō: Weatherhill.

Ohmae Kenichi, ed. 1983. *Japan: Obstacles and Opportunities, (A Binational Perspective for U.S. Decision-makers Prepared by McKinsey & Company, Inc. for the United States-Japan Trade Study Group* Tōkyō: President Inc.

Ouchi, William G. 1981. *Theory Z: How American Business Can Meet the Japanese Challenge*. Reading, Massachusetts: Addison-Wesley.

Panko, Raymond R. 1983. "Kanji Keyboard Chaos." *Computerworld*, November 28.

Salthouse, Timothy A. 1984. "The Skill of Typing." *Scientific American*, February, 128-135.

Suzuki Takao. 1977. "Writing Is Not Language, or Is It?" *Journal of Pragmatics*, 1:4 (December) 407-420.

Tōdō Akiyasu. 1982. *Kanji no kako to mirai* [The past and future of Chinese characters]. Iwanami shinsho 205. Iwanami shoten.

Vogel, Ezra F. 1979. *Japan As Number One: Lessons for Americans*. Cambridge, Massachusetts & London: Harvard University Press.

Weizenbaum, Joseph. 1976. *Computer Power and Human Reason: From Judgment to Calculation*. San Francisco: W. H. Freeman.

Yamada Hisao. 1980. "A Historical Study of Typewriters and Typing Methods from the Position of Planning Japanese Parallels." *Journal Information Processing* (Tōkyō) 2:5.175-202.