The Student-Computer Interface of an Intelligent Tutoring System for Japanese Language Instruction

Anthony A. Maciejewski
School of Electrical Engineering
Purdue University
West Lafayette, IN 47907

Abstract

This work describes the human-computer interaction of an intelligent tutoring system designed to mediate some of the difficulties of acquiring proficiency in reading technical Japanese material.

I. Introduction

In recent years, Japan has emerged as an acknowledged leader in several key areas of applied science and engineering. Japanese funding of R & D on a per capita basis equals that of U.S. civilian R & D expenditures [1] and it's output of technical publications has steadily grown to 8% of the world total, second only to the United States [2]. Unfortunately, access to this wealth of information is denied to most Western scientists and engineers—of the approximately 10,000 technical journals that are published by the Japanese, only an estimated 6% are available in English [1].

The lack of Japanese language proficiency among American scientists and engineers is due to a number of factors including the inherent difficulty of the language and the scarcity of technical curricula and course materials [3]. The goal of this work is to mitigate some of these difficulties through the use of an intelligent computer tutoring system. The remainder of this paper first presents a brief summary of some of the unique characteristics of the Japanese language and is followed by a description of the student-computer interface of an intelligent tutoring system developed to accelerate the acquisition of technical Japanese reading proficiency.

II. The Japanese Language

The Japanese language is generally regarded as one of the most difficult languages for English-speaking people to learn. While the number of individuals studying Japanese is increasing there remains an extremely high attrition rate, estimated by some to be as high as 80% [3]. Much of this difficulty can be associated with the Japanese writing system. Japanese text consists of two distinct orthographies, a phonetic syllabary known as kana and a set of logographic characters, originally derived from the Chinese, known as kanji. The kana are divided into two phonetically equivalent but graphically distinct sets, katakana and hiragana, both consisting of 46 symbols and two diacritic marks denoting changes in pronunciation. The katakana are used primarily for writing words of foreign origin that have been adapted to the Japanese phonetic system although they are also used for onomatopoeia, colloquialisms and emphasis. The hiragana are used to write all inflectional endings and some types of native Japanese words that are not currently represented by kanji. Due to the limited number of kana, their relatively low visual complexity, and their systematic arrangement they do not represent a significant barrier to the student of Japanese. In fact, the relatively small effort required to learn katakana yields significant returns to readers of technical Japanese due to the high incidence of terms derived from English and transliterated into katakana.

In contrast, the ability to read the much more visually complex kanji, which are used to write the vast majority of words functioning as nouns, verbs, or adjectives, presents a formidable task, even for native Japanese. A dictionary of all kanji ever used would contain on the order of 50,000 entries. Fortunately, 90% of all the currently used kanji can be found in the six to seven thousand entries specified under the Japanese Standard Association's JIS X 0208 standard. Though frequently considered to be graphical representations of objects or concepts, in actuality very few kanji belong to the category of pictographs or ideographs. The vast majority (over 90%) can be classified as phonologograms, characters composed from the combination of a phonetic component and a general meaning component. The phonetic component, however, does not uniquely specify the pronunciation of a character. This is due to the fact that kanji have multiple readings depending on the context. These readings are generally classified as either native Japanese (kun) or as being derived from the Chinese (on). The kun readings are generally used for single-character kanji or with inflected forms whereas on readings are more frequently used in multiple kanji words. Exceptions abound and phonologograms are frequently used for their phonetic component while disregarding the meaning or vice versa (such compounds are called oseji).

The difficulty of reading such a diverse and complicated character set is further magnified by the fact that Japanese text does not possess unique characters to specify lexical boundaries that correspond to the spaces used between words in Western languages. Thus a student has no idea whether a single kanji is to be interpreted as a com-
plete word or whether it is part of a compound. Likewise with kanji one must know whether it is being used to repre-
sent an inflectional ending or a separate word. Additional
difficulty is presented by the fact that Japanese sentence
structure is basically Subject - Object - Verb (SOV) as
opposed to the more common SVO structure of English.

III. Student-Computer Interaction

Many of the aspects of written Japanese discussed
above present serious difficulties to the student, partic-
ularly with respect to the use of traditional study aids such
as dictionaries and printed reference materials. Much of
this is due to the fact that a significant amount of infor-
mation must already be known about a Japanese "word"
(number of characters, pronunciation, stroke counts, etc.)
before it can even be accessed in such works. Thus even
the use of phonetic conversion software of the type typi-
cally used in Japanese word processors [4-8] provides min-
imal assistance for identifying unknown passages of text.
The situation is further complicated by not relying on any
categorization scheme based on phonetic alphabetization
or the pattern of strokes used to create the character. The
interaction of the tutor with a student is presented in the
following example of a typical session.

The tutor first addresses the issue of context by iden-
tifying the student and accessing their personal database
which contains information about the their technical inter-
est as well as language proficiency. The keywords which
define the student's area are used to search the electronic
database of technical Japanese material. This database
currently consists only of selected Japanese abstracts and
their associated English translations. The database
of language proficiency consists of the kanji and kanji com-
 pounds, along with their specific readings, to which the
student has been exposed. These are divided into those
which have been mastered and those which require further
review. The database also includes a list of grammatical
structures with which the student is familiar. The text
which is chosen by the tutor for the student is thus de-
signed to reinforce material which is not yet mastered and
to introduce new material within a specific context. The
order in which new kanji are presented is based on fre-
cquency counts within the desired discipline.

The tutor presents the actual Japanese text on the
terminal screen in exactly the same format as it appears
in the original Japanese publication. The addition of pho-
netic readings in kanji, known as furigana, is avoided al-
though this practice is common in language textbooks.
This reduces the tendency of the student to rely on the
furigana while reading thus preventing assimilation of the
kanji. The text, while designed to maximize comprehen-
sion from context, will clearly have material which is not
grouped by the student. When this occurs the student can
highlight the difficult passage with a mouse and request a
translation from this text. The tutor, detecting an input from
the student, takes the highlighted Japanese text, matches it with it's English translation, and displays
this on the screen. At this point it takes all of the vocabu-
lar and grammatical structures present in the highlighted
passage and updates the student's personal database to
include these items in the material that is not currently
mastered. Thus those items with which the student has
difficulty will be repeated with greater frequency in subse-
quent sessions.

The manner in which the student highlights the un-
known text is significant with respect to the information
stored by the tutor. Any amount of text can be highlighted
ranging from a single character to an entire sentence. Thus
the tutor attempts to pinpoint the exact nature of the dif-
ficulty by determining whether the highlighted region is
an inflectional ending, a single word, a clause, etc. There
is particular interest in analyzing student responses that
result in regions which do not occur at morphological or
lexical boundaries. This is, of course, a common occur-
dence due to the lack of spacing between Japanese words.

In the case that the student actually wants information on
the sentence structure, the interface is designed to allow
the student to click on the mouse while in a highlighted
region in order to move one level up in the parse tree rep-
resenting the sentence. Thus the student can trace the
structure from a single morpheme to the entire sentence.

In addition to the above general scenario of a tutoring
session, the tutor is being outfitted with electronic versions
of supplementary materials such as examples of handwritten
text and grammar references as well as general, techni-
cal, and character dictionaries. These are all linked to the
graphical interface so that immediate access to all entries
related to the highlighted region can be obtained. This
environment should accelerate the rate at which scientists
and engineers can acquire a reading knowledge of technical
Japanese.

This material is based upon work supported by the
National Science Foundation under Grant No. INT-
8818039. The Government has certain rights in this
material.

References

[1] P. H. Hubbard, President, Electronics Education
Foundation, American Electronics Association, Test-
imony before House Subcommittee on Science, Re-

U.S. Engineering Through International Coopera-
tion: Some Recommendations for Action, Washington,

Technical Japanese for Scientists and Engineers: Cur-
ricular Options, a report to the National Science
Foundation, Massachusetts Institute of Technology,

teific American, pp. 96–107, July 1984.

orean," IEEE Computer, Vol. 18, No. 1, pp. 27–34, Jan-
uary 1985.