

A Student Model of *Katakana* Reading Proficiency for a Japanese Language Intelligent Tutoring System

Anthony A. Maciejewski Yun-Sun Kang

School of Electrical Engineering
Purdue University
West Lafayette, Indiana 47907

Abstract—This work describes the development of a student model that is used in a Japanese language intelligent tutoring system to assess a pupil's proficiency at reading one of the distinct orthographies of Japanese, known as *katakana*. While the effort required to memorize the relatively few *katakana* symbols and their associated pronunciations is not prohibitive, a major difficulty in reading *katakana* is associated with the phonetic modifications which occur when English words which are transliterated into *katakana* are made to conform to the more restrictive rules of Japanese phonology. The algorithm described here is able to automatically acquire a knowledge base of these phonological transformation rules, use them to assess a student's proficiency, and then appropriately individualize the students' instruction.

I. INTRODUCTION

Interest in Japanese language instruction has risen dramatically in recent years, particularly for those Americans engaged in technical disciplines. However, 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% [7]. 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

Manuscript received July 29, 1991. This material is based upon work supported by the National Science Foundation under Grant No. INT-8818039 and in part by a grant from NEC Corporation.

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*.

This work describes the development of a system that is used to automatically acquire knowledge about how English words are transliterated into *katakana* and then to use that information in developing a model of a student's proficiency in reading *katakana*. This model is used to guide the instruction of the student using an intelligent tutoring system developed previously [6, 8]. The remainder of this paper is organized as follows: Section II provides a brief introduction to the *katakana* writing system and to the rules of Japanese phonology. This is followed by the description of an algorithm to automatically generate a *katakana* to English dictionary from an arbitrary Japanese text and its English translation. In section IV, the *katakana* to English dictionary is used to develop a set of phonological rules that govern the inverse transformation from Japanese phonetics, represented by the *katakana* orthography, back to the original English pronunciation. This set of rules is treated as the knowledge base that the student must acquire in order to become proficient at reading *katakana*. The failure to recognize words that contain specific phonological rules is then used to build the student model, which is described in section V. Finally, the conclusions of this work are presented in the last section.

II. KATAKANA AND JAPANESE PHONOLOGY

The Japanese lexicon contains an extremely large number of words originating from foreign languages. Traditionally, these Japanese loan words are characterized as either *kango*, literally "Chinese word," which are of Chinese origin (also referred to as Sino-Japanese words) or *gairaigo*, literally "foreign word," for words from any other country of origin. While the proportion of Sino-Japanese words in the lexicon is extremely large due to the profound cultural influence of China, words of English origin have dominated the class of loan words since the late 19th century. In a study of Japanese publications performed between 1956 and 1964, over 80% of the *gairaigo* originated from English [4]. This process of adopting English

into the lexicon is particularly common for relatively new or specialized terms arising in technical literature.

When adopting a word of foreign origin into Japanese, the original pronunciation of that word is typically transliterated into *katakana* which graphically represents all of the possible phonetic sequences in the Japanese language. It is this process of modifying English phonetic sequences to conform to the rules of Japanese phonology which presents English-speaking readers of *katakana* with difficulty in identifying a word's meaning. This is due to the fact that the rules of Japanese phonology are quite different from those of English. In particular, Japanese has only five single vowels /a i u e o/ in contrast to the large number of vowels in English. These vowels, when combined with the nine Japanese consonants /k s t n h m j r w/ constitute 44 of the 46 basic sounds in Japanese which are traditionally organized as shown in Table I. This organization is referred to as *gozyūōnzu*, which literally translates as "table of fifty sounds," since all fifty sounds were used at one point in time. The pronunciation of each *katakana* character in the table is represented using the International Phonetic Alphabet (IPA) symbols. In addition to these basic pronunciations, certain *katakana* may be written with one of two diacritic marks that modify their pronunciation (see Table II). In particular, the four rows in Table I that correspond to the consonants /k s t h/ may be written with a symbol consisting of two short parallel lines, known as *nigori*, which results in a voiced version of these consonants, i.e. /g z d b/. The second diacritic mark is a small circle referred to as *maru* which when combined with the *katakana* in the /h/ row results in pronunciations which include the bilabial stop /p/. The *katakana* in the second column, i.e. those associated with the vowel /i/, may also be combined with any of the *katakana* in the row associated with the consonant /j/ to produce additional single syllable pronunciations. The second *katakana* in this case is written slightly smaller in order to differentiate from a two syllable sequence, as illustrated in Table III. Finally, Japanese includes two moraic consonants, /Q/ the mora obstruent and /N/ the mora nasal, which are presented in Table IV.

From the above description of the *katakana* orthography, it is clear that certain inherent limitations are imposed on phonetic sequences in the Japanese language. In particular, Japanese does not allow any consonant clusters except when a consonant is followed by a glide or preceded by a moraic consonant [10]. In addition, consonants may not appear at the end of a sequence. These restrictions, together with the limited number of Japanese vowels, result in the vast majority of phonological modifications which occur when transliterating an English word into *katakana*. By the same token, these resulting modifications are the source of difficulty for English-speaking readers of *katakana*.

It should also be noted that there are additional difficulties to comprehending *katakana* unrelated to the phonological processes involved. In particular, while it is true that the vast majority of loan words are created by the phonological process, foreign borrowings may also be modified by changes in form due to simplification, semantics, or Japanese coinage [9]. For example, simplification frequently occurs with

TABLE I
KATAKANA CHARACTERS AND THEIR PHONETIC REPRESENTATION
IN IPA SYMBOLS: BASIC SYLLABLES

ア	イ	ウ	エ	オ
/a/	/i/	/u/	/e/	/o/
カ	キ	ク	ケ	コ
/ka/	/ki/	/ku/	/ke/	/ko/
サ	シ	ス	セ	ソ
/sa/	/ʃi/	/su/	/se/	/so/
タ	チ	ツ	テ	ト
/ta/	/tʃi/	/tsu/	/te/	/to/
ナ	ニ	ヌ	ネ	ノ
/na/	/ni/	/nu/	/ne/	/no/
ハ	ヒ	フ	ヘ	ホ
/ha/	/hi/	/fu/	/he/	/ho/
マ	ミ	ム	メ	モ
/ma/	/mi/	/mu/	/me/	/mo/
ヤ		ユ		ヨ
/ja/		/ju/		/jo/
ラ	リ	ル	レ	ロ
/ra/	/ri/	/ru/	/re/	/ro/
ワ				
/wa/				

TABLE II
KATAKANA CHARACTERS AND THEIR PHONETIC REPRESENTATION
IN IPA SYMBOLS: MODIFIED SYLLABLES

ガ	ギ	グ	ゲ	ゴ
/ga/	/gi/	/gu/	/ge/	/go/
ザ	ジ	ズ	ゼ	ゾ
/za/	/dʒi/	/zu/	/ze/	/zo/
ダ	ヂ	ヅ	デ	ド
/da/	/dʒi/	/zu/	/de/	/do/
パ	ピ	プ	ペ	ポ
/pa/	/pi/	/pu/	/pe/	/po/
バ	ビ	ブ	ベ	ボ
/pa/	/pi/	/pu/	/pe/	/po/

TABLE III
KATAKANA CHARACTERS AND THEIR PHONETIC REPRESENTATION
IN IPA SYMBOLS: CONSONANT PLUS /j/, /jʲ/, or /jʲo/

キヤ	キユ	キョ	ギヤ	ギユ	ギョ
/kja/	/kju/	/kjo/	/gja/	/gju/	/gjo/
シヤ	シユ	ショ	ジヤ	ジユ	ジョ
/ʃja/	/ʃju/	/ʃjo/	/dʒja/	/dʒju/	/dʒjo/
チャ	チュ	チョ			
/tʃja/	/tʃju/	/tʃjo/			
ニヤ	ニユ	ニョ			
/nja/	/nju/	/no/	ビヤ	ビユ	ビョ
ヒヤ	ヒユ	ヒョ	/bjja/	/bjju/	/bjjo/
/hja/	/hju/	/ho/	ピヤ	ピユ	ピョ
ミヤ	ミユ	ミョ	/pjja/	/pjju/	/pjjo/
/mja/	/mju/	/mo/			
リヤ	リュ	リョ			
/lja/	/lju/	/lo/			

TABLE IV
KATAKANA CHARACTERS AND THEIR PHONETIC REPRESENTATION
IN IPA SYMBOLS: MORAL CONSONANTS

ン	ッ
/N/	/Q/

polysyllabic words such as “television” and “word processor” which are shortened to the *katakana* words *tere* and *waapuro*, respectively. Changes in semantics have resulted in the *katakana* word *botan* being used to designate a touch tone type of telephone, whereas its phonetic origin is from the word button. Examples of coinage which result from combinations of existing loan words include *maikaa* (derived from my + car) and *mairoomu* (derived from my + car) which refer to privately owned cars and houses. All of these processes contribute to a student's difficulty in achieving reading proficiency in *katakana*, however, this work focuses on the phonological modifications.

III. A KATAKANA TO ENGLISH DICTIONARY

Due to the relatively specialized and technical nature of a significant portion of the recently borrowed loan words, comprehensive *katakana* to English dictionaries are not readily available. Therefore, the first step in analyzing the phonetic modifications which occur when English words are transliterated into *katakana* is to develop an algorithm for automatically generating a *katakana* to English dictionary from Japanese text for which English translations are available. Fortunately, many examples of Japanese technical literature which contain significant amounts of *katakana* are available in an electronic format, along with the English translation of the document. The primary source for the development of the *katakana* to English dictionary described here, is a set of 3,000 Japanese phrases from the telecommunications thesaurus obtained courtesy of NTT (Nippon Telephone and Telegraph). Of the 3,000 phrases, approximately one half of them contain at least one example of a word written in *katakana*.

The *katakana* words in an electronic Japanese document are easily identified by their unique values within the JIS coding system, the Japanese version of ASCII. The problem then becomes one of identifying the English word from which the *katakana* word was derived. Fortunately, there is typically a sentence to sentence correspondence between Japanese text and its English translation so that the number of possible English candidate words is restricted to a single sentence. The first step in identifying the corresponding English word or words, is to first convert the *katakana* sequence into *roomaji*, i.e., the traditional English letters. While several variations of transliterating *katakana* into English exist, the method used here is known as the Hepburn system. A table of *katakana* to English transliteration rules for the Hepburn system is presented in Table V.

Once the Japanese *katakana* has been transliterated into English characters, the process of correlating this text string with the English text from which it originated can be simply considered as a string pattern matching problem. Since the transliterated *katakana* will not typically match the spelling of the English text from which it originated, an approximate string matching algorithm based on dynamic programming is employed [11]. Thus the *katakana* word that has been transliterated into English ASCII characters is considered as the pattern for which one would like to find the best

TABLE V
THE MODIFIED HEPBURN ROMANIZATION SYSTEM

a	i	u	e	o
ka	ki	ku	ke	ko
sa	shi	su	se	so
ta	chi	tsu	te	to
na	ni	nu	ne	no
ha	hi	hu	he	ho
ma	mi	mu	me	mo
ya		yu		yo
ra	ri	ru	re	ro
wa				
ga	gi	gu	ge	go
za	ji	zu	ze	zo
da	ji	zu	de	do
ba	bi	bu	be	bo
pa	pi	pu	pe	po
kya		kyu		kyo
sha		shu		sho
cha		chu		cho
nya		nyu		nyo
hya		hyu		hyo
mya		myu		myo
rya		ryu		ryo
gya		gyu		gyo
ja		ju		jo
bya		byu		byo
pya		pyu		pyo
	n̄			

approximate match in the corresponding English sentence. Presumably, this should be the English word or words from which the *katakana* was derived. Unfortunately, two factors prevent this straightforward approach from performing satisfactorily. The first is due to the fact that the original transliteration from English to *katakana* results in a significant change in pronunciation and the second is due to the highly irregular spelling conventions in English. To compensate for the first factor, two techniques are employed. The first is to apply a set of possible spelling changes to the transliterated *katakana* that attempt to account for some of the major differences between Japanese and English phonology. These possible spelling changes are listed in Table VI and roughly correspond to the suggestions given to students of Japanese when trying to decipher *katakana* [3]. The second technique used to improve the pattern matching process is a nonuniform weighting on the ASCII symbols, with more importance placed on the matching of consonants. This is motivated by the poor correspondence between the five vowel sounds in Japanese with the numerous vowels in English, as well as by the large variations in spelling.

With the modification of the approximate pattern matching algorithm to include the consonant weighting and spelling change rules, the *katakana* that appear in the sample of 1,500 telecommunications phrases can be matched to their correct corresponding English translation with 90% accuracy. The remaining errors are primarily a result of the irregularities in English spelling. This effect can be removed by performing the string matching based on the pronunciation of the strings rather than their spelling. When both the *katakana* and the English text are converted to strings of phonemes, the error

TABLE VI
LIST OF SPELLING RULES USED AFTER TRANSLITERATION OF
KATAKANA INTO ENGLISH USING THE HEPBURN SYSTEM

NO	RULE	katakana word	English origin
1	$u \rightarrow * / \% _ (\% \#)$	<i>shisutemu</i>	system
2	$o \rightarrow * / (d t) _ (\% \#)$	<i>doraiba</i>	driver
3	$i \rightarrow * / (z \%) _ (\% \#)$	<i>matchi</i>	match
4	$howa \rightarrow wh / _ \$$	<i>howaito</i>	white
5	$(u uu) \rightarrow w / _ \$$	<i>uuru</i>	wool
6	$i \rightarrow y$	<i>iesu</i>	yes
7	$ee \rightarrow y\$$	<i>eeru</i>	Yale
8	$y \rightarrow * / _ \$$	<i>kyaburetaa</i>	carburetor
9	$a \rightarrow \$r / (oo o e i) _ \#$	<i>hea</i>	hair
10	$aa \rightarrow \$r$	<i>misutaa</i>	mister
11	$a \rightarrow \$r / _ \#$	<i>konpyuta</i>	computer
12	$oo \rightarrow \$r$	<i>pooku</i>	pork
13	$s \rightarrow c / _ e$	<i>sero</i>	cello
14	$s \rightarrow th$	<i>sumisu</i>	Smith
15	$z \rightarrow j / _ e$	<i>zerii</i>	jelly
16	$z \rightarrow th$	<i>mazaa</i>	mother
17	$j \rightarrow (d z)$	<i>ejison</i>	Edison
18	$b \rightarrow v$	<i>banira</i>	vanilla
19	$h \rightarrow f$	<i>haaiisuto</i>	Far East
20	$r \rightarrow l$	<i>rengusu</i>	length
21	$ts \rightarrow (t z)$	<i>tsurii</i>	tree

The symbols “%”, “\$”, and “#” represent consonant, vowel, and sequence-boundary, respectively.

rate is reduced to less than 0.1%.

IV. PHONETIC RULES FOR KATAKANA CONVERSION

The *katakana* to English dictionary generated in the previous section can now be used to determine the specific phonetic changes that occur in English text when it is transliterated into *katakana*. This section describes an algorithm for automatically determining a set of rules that govern these phonetic changes. The inputs into the algorithm are a string of *katakana* and the English text from which it was derived. These two inputs are then converted into IPA symbols using Tables I thru IV for the *katakana* and the UNIX version of Webster's *Seventh New Collegiate Dictionary* [2] for the English.

The phonological rules that are determined by the algorithm presented here consist of three elements: a source phoneme, a target phoneme, and conditions. The source phonemes are those that are derived from the *katakana* string (since this is the text that the student will be reading) and the target phonemes are those that are derived from the English text. The following notation, adapted from [5], is used to represent a phonological rule:

$$S \rightarrow T / C_1 _ C_2$$

where S is a phoneme of the source language, T is a phoneme

of the target language, and C_1 and C_2 are the pre-condition and post-condition, respectively, that must be satisfied for S to be replaced by T. The conditions C_1 and C_2 are either single phonemes or phonological categories such as vowels, consonants, etc.

Phonological rules of the type described above cannot be directly generated from the IPA symbols of the raw input strings because there is not a correspondence between the *i*th phoneme of the source string and the *i*th phoneme of the target string. This is primarily due to the different constraints on consonant clusters in Japanese as compared to English. As mentioned above, Japanese does not, in general, allow consonant clusters whereas English allows initial consonant clusters of two or three consonants, as in such words as “sky” and “spray”, and either two, three, or four final consonants, as in “ask”, “elks”, and “glimpsed” [1]. To compensate for these phonological differences, the sequence of IPA symbols for the English text string is modified by inserting the symbol “*”, which has no pronunciation, between any consecutive consonant phonemes and after sequences that end with a consonant. This modification greatly improves the correspondence between phoneme symbols in the *katakana* string with the symbols in the English string. However, it does create a difficulty with the mora nasal consonant cluster that is allowed in Japanese. It has been observed that the mora nasal /N/ when followed by a consonant corresponds to either /m/, /n/, or /ŋ/ followed by a consonant in English. Therefore, in order to provide uniform treatment of these consonant clusters, the symbol “*” is also inserted into the *katakana* phoneme sequence between the mora nasal and the following consonant.

In addition to consonant clusters, one must also consider how to deal with non-identical vowel sequences in order to improve the correspondence between symbols in the source string and the target string. In Japanese, it is not obvious when to consider the second vowel of a sequence as a separate syllable as opposed to the second half of a diphthong. This creates a difficulty when trying to match the English vowels. Therefore, the approach adopted here is that all consecutive vowel sequences are treated as a single unit. This is true even if it is known that the English vowel sequence represents two separate syllables. Therefore, after the appropriate modification of the input strings by inserting “*” symbols, the strings are divided into units which consist of either a single consonant, a sequence of vowels or semi-vowels, or the symbol “*”. If the number of partitions in the two strings are equal, then a set of phonological rules is generated.

The algorithm described above can be summarized by the following four steps:

- (1) Convert the *katakana* input string and the English input string into IPA symbols.
- (2) Modify the English IPA string by inserting a “*” symbol between any consecutive consonant phonemes and after a final consonant. Modify the Japanese IPA string by inserting a “*” symbol between the mora nasal /N/ and any following consonant.
- (3) Divide both symbol strings into partitions which

TABLE VII
LIST OF PHONOLOGICAL RULES CREATED BY MATCHING THE IPA EQUIVALENT OF A KATAKANA WORD WITH THAT OF ITS ENGLISH ORIGIN

Rule		Data		Example			
Jap	Eng	Probability (%)	Number of Occurrences	カタカナword	English word	IPA (JAP)	IPA (ENG)
*	w	63	5	シーケンス	sequence	ʃi-k(*)ensu	sikwens
a	ə	36	156	デジタル	digital	didʒitaru	dɪdʒətəl
a	ə	24	103	コンピュータ	computer	koNpju-ta	kəmpjutə
a	æ	34	145	セラミック	ceramic	seramiQku	səɾæmɪk
a-	ə	69	43	サーマル	thermal	sa-maru	θəməɪ
b	v	34	44	バルブ	valve	barubu	vælv
dʒ	z	19	10	ビジー	busy	bidʒi-	bɪzi
e	ə	28	58	ドキュメント	document	dokjumeNto	dəkjəmənt
h	f	21	8	ヒューズ	fuse	hju-zu	fjuz
i	ə	18	63	サービス	service	sa-bisu	səvəs
i	ɪ	48	165	レジスタ	resistor	redʒisuta	rɪzɪstə
o	*	50	208	キーボード	keyboard	ki-bo-do	kɪbɔrd(*)
o	ə	20	81	セッション	session	seʃjoN	seʃən
o	a	15	64	プロセス	process	puɾosesu	pɾəs
r	l	57	270	ライン	line	rain	laɪn
s	θ	4	11	レンクス	length	reNɡusu	leŋθ
ʃ	s	47	33	システム	system	ʃisutemu	sɪstəm
tʃ	t	42	14	チューブ	tube	tʃu-bu	tʃub
u	*	91	609	チェック	check	tʃeQku	tʃek(*)

include either a single consonant, a consecutive sequence of vowels and/or semi-vowels, or the symbol “*”.

(4) If the number of partitions in the source string is equal to that of target string, generate a phonological rule which defines the transformation of a source partition into a target partition.

This process is illustrated in the following example, in which the *katakana* word *shisutemu* is compared with the word “system” from which it was derived:

システム → /ʃisutemu/ → /ʃ,i,s,u,t,e,m,u/
system → /sɪstəm/ → /s,ɪ,s,*,t,ə,m,*/.

The following eight rules are generated:

ʃ → s / # _i
i → ɪ / ʃ _s
s → s / i _u
u → * / s _t
t → t / u _e
e → ə / t _m
m → m / e _u
u → * / m _#.

Note that the somewhat obvious rules in which a phoneme is not changed are significant since the same phoneme may be modified in a different context. Additional information regarding the relative probability of occurrence for rules which

have identical conditions is also maintained.

V. STUDENT MODEL

The phonological rules determined in the previous section are used to define the domain knowledge that a student must acquire in order to become proficient in reading *katakana*. Some examples of these rules resulting from processing the 1,500 entries in the *katakana* to English telecommunications dictionary are illustrated in Table VII. For each student that uses the intelligent tutoring system, a database is maintained which keeps information on which of these rules the student has mastered and which ones need further review. This information is obtained both by passive monitoring of the student's requests for information while using the tutoring system [6] as well as through active testing of the student. The *katakana* words with which a student has difficulty are analyzed to determine which phonological rules are present and this is compared to the rules which occur in *katakana* that are easily comprehended. The tutoring system then attempts to assist the student in acquiring the unfamiliar rules by presenting lessons which include *katakana* that use these rules.

The following presents a simple illustration of how the student model is formed by analyzing the responses of a student who was tested for his *katakana* reading proficiency. This particular student had no difficulty with the following words:

<i>shisutemu</i>	→	system
<i>bideo</i>	→	video
<i>totaru</i>	→	total
<i>tasuku</i>	→	task

which include the rules $u \rightarrow *$, $f \rightarrow s$, $i \rightarrow I$, $b \rightarrow v$, and $r \rightarrow l$. However, the student could not comprehend the following *katakana* words:

<i>aasu</i>	→	earth
<i>rengusu</i>	→	length
<i>saamaru</i>	→	thermal

which use the phonological transformation rules: $u \rightarrow *$, $r \rightarrow l$, and $s \rightarrow \theta$. From analyzing these two sets of data, the tutoring system is able to correctly identify that the rule which the student has not mastered is the transformation $s \rightarrow \theta$. This is not particularly surprising since this is a rather radical change in pronunciation which occurs relatively infrequently (see Table VII). Indeed, this student is rather typical in that he has acquired the relatively straightforward rules such as $u \rightarrow *$ which occurs extremely frequently and is one of the primary mechanisms for dealing with the disparity in consonant clusters between English and Japanese. Likewise, this student has no trouble with the simple consonant substitutions $b \rightarrow v$ or $r \rightarrow l$. Therefore, the tutoring system would tailor the instruction of this student with *katakana* words that contain the more obscure rules such as $s \rightarrow \theta$, hopefully being able to find occurrences in which this is the only rule present in order to provide more contextual information.

VI. CONCLUSIONS

The goal of this work was the development of a model for representing a student's proficiency in reading *katakana*. This model is used to individualize the instruction of an intelligent tutoring system that is designed to assist scientists and

engineers acquire a reading knowledge of technical Japanese. To accomplish this goal an algorithm was first developed to automatically generate a *katakana* to English dictionary from raw Japanese text and its English translation. This dictionary is then used as input into a second algorithm that is used to generate the phonological rules that govern the transformations that English words undergo when being transliterated into *katakana*. These rules are used as the domain knowledge base against which a student's performance is measured. This technique has proven very effective in analyzing the difficulties which students encounter in comprehending *katakana*.

REFERENCES

- [1] J. C. Catford, *A Practical Introduction to Phonetics*. New York, NY: Oxford University Press, 1988.
- [2] P. B. Gove, ed., *Webster's Seventh New Collegiate Dictionary*. Springfield, MA: G. & C. Meriam Company, 1963.
- [3] E. H. Jordan and H. I. Chaplin, *Reading Japanese*. New Haven and London: Yale University Press, 1976.
- [4] Kokuritsu Kokugo Kenkyuujo, *Gendai-zasshi 90shu no yoogo yooji*. vol. 3, Report 25, 1964.
- [5] K. Kostenniemi, "Two level model for morphological analysis," *The 8th Int. Joint Conf. Artificial Intelligence*, pp. 683-685, 1983.
- [6] N. K. Leung and A. A. Maciejewski, "The Nihongo Tutorial System for learning technical Japanese," *1991 Nat. Educ. Comp. Conf.*, pp. 143-148, Phoenix, AZ, June 18-20, 1991.
- [7] D. O. Mills, R. J. Samuels, and S. L. Sherwood, *Technical Japanese for Scientists and Engineers: Curricular Options*, a report to the National Science Foundation, Massachusetts Institute of Technology, MITJSTP WP 88-02, 1988.
- [8] A. A. Maciejewski, "The student-computer interface of an intelligent tutoring system for Japanese language instruction," *Proc. 1989 IEEE Int. Conf. Sys., Man, Cyber.*, pp. 311-312, Cambridge, MA, Nov. 14-17, 1989.
- [9] M. Shibata, *The Languages of Japan*. New York, NY: Cambridge University Press, 1990.
- [10] J. T. Vance, *An Introduction to Japanese Phonology*. Albany, NY: State University of New York Press, 1987.
- [11] R. A. Wagner and M. J. Fischer, "The string-to-string correction problem," *J. ACM*, vol. 21, no. 1, pp. 168-178, Jan., 1974.