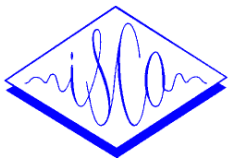
ISCA Archive
http://www.isca-speech.org/archive



First International Conference on
Spoken Language Processing (ICSLP 90)
Kobe, Japan
November 18-22, 1990

THE LINGUISTIC PROCESSING MODULE FOR JAPANESE TEXT-TO-SPEECH SYSTEM

Tohru Shimizu, Norio Higuchi, Hisashi Kawai and Seiichi Yamamoto

KDD R&D Laboratories
Kamifukuoka-shi, Saitama, 356 Japan

ABSTRACT

A linguistic processing module for Japanese text-to-speech system for use on personal computers has been developed. The linguistic processing module analyzes Japanese sentences and outputs information for the rule-based speech synthesizer. This module is implemented as software for 16-bit personal computers, running under the operating system MS-DOS.

The linguistic processing module has the following two characteristics.

(1) Both a morphological analysis module and a *Kana-Kanji* conversion module are available to analyze input texts, and the editing module can modify the output of these two modules.

(2) The syntactic analysis which is introduced in morphological analysis decreases the number of word candidates, and consequently the system has high performance for selection of correct words.

The result of the evaluation tests shows that the accuracy of the morphological analysis is 99.0% for the sample articles of a popular economical newspaper.

I. INTRODUCTION

Any sentence can be converted to synthetic speech using a rule-based speech synthesizer. Recently, text-to-speech systems have been put to practical use [1]. The authors developed the text-to-speech system with word processing functions three years ago. The system used a special word processor, which had the functions for speech synthesis use. For example, (1) pronunciation and prosodic symbols were determined when input *Kana* is converted to *Kanji-Kana* mixed sentence, (2) the properties of the speech can be selected by the menu on the screen, (3) words can be speech output by the operator, by selecting one candidate from several word candidates, (4) the speech can be transmitted through the telephone network [2]. The combination of the word processor and the rule-based speech synthesizer provided a versatile environment to output high quality speech easily. Though the system could output any sentence, the text had to be created by this word processor because this system used the special word processor for linguistic processing module. In order to convert texts which are already made with other word processors, the morphological analysis module has been developed.

This paper describes the configuration of the system, the linguistic processing techniques employed in morphological analysis, and the results of the evaluation tests.

II. CONFIGURATION OF THE SYSTEM

The configuration of the system is shown in Fig. 1. The linguistic processing software of this system consists of four modules: the morphological analysis module, the *Kana-Kanji* conversion module, the text editing module(word processor), and the prosodic control module.

The morphological analysis module and the *Kana-Kanji* conversion module determine the position of word boundaries for the *Kanji-Kana* mixed sentences and *the Kana* sentences respectively. The operator can choose either module, since both use a common dictionary and output data format.

The *Kana-Kanji* conversion module has a special function to help an operator select the correct word candidate. This system displays a part of speech for each word candidate as well as an expression itself, because the accent type of the word to be synthesized is determined by the expression and the part of speech in many cases. The operator can select the correct word candidate based on the expression and the part of speech. Not only *Kana-Kanji* conversion but also morphological analysis may cause errors, therefore the *Kana-Kanji* conversion module is also used to correct errors in the result of the morphological analysis.

The output of each module is stored in the *Bunsetsu*-file. In Japanese, a *Bunsetsu* consists of a content word and zero or more function words. The editing module allows the operator to modify the *Bunsetsu*-file, and add attributes of a speech sound (speaker, speed, etc.).

The prosodic control module determines the intonation and accentuation based on the information of the *Bunsetsu*-file. This module outputs alphabet strings with prosodic symbols to the rule-based speech synthesizer.

III. LINGUISTIC PROCESSING TECHNIQUES OF MORPHOLOGICAL ANALYSIS

This section describes the linguistic processing technique employed in the morphological analysis module.

As Japanese sentences are not written with a space between the words, the word boundary must be determined before syntactic analysis. The position of word boundaries are determined by testing all possible combinations of a content word and following function words included in the content word dictionary and the function word dictionary.

As all proper nouns are not included in the dictionary, undefined words are assumed. The part of speech and accent type of undefined words are estimated in this process.

The combination of the words is made after the word boundaries are determined. The order of the priority is given for each combination of words to decrease the number of word candidates. After that one combination of words is selected by using syntactic information.

The block diagram of the morphological analysis module is shown in Fig. 2.

3.1 Content word dictionary

The content word dictionary has approximately 50,000 words. Each entry word has a part of speech, accent type and syntactic information. The word is treated as the different word when at least one attribute is different.

(1) treated as the different word

same meaning and different pronunciation

山 (mountain) pronunciation: yama,san

浅間-山 asama-yama

阿蘇-山 aso-san

same part of speech and different accent type

もう (more, already) part of speech: adverb

accent type: more 0

already 1

(2) treated as the same word

different meaning but the same part of speech and accent

山 (mountain, a pile) part of speech: noun

accent type: 0

3.2 Assumption of an undefined word

The assumption of an undefined proper noun is very important to increase the accuracy of the morphological analysis because the content word dictionary doesn't include all proper nouns.

Before the selection of the words, undefined words are assumed. The part of speech and accent type are determined when the word boundaries of undefined word are assumed because the part of speech and accent type are very important in pronouncing the proper noun correctly. This system includes single characters in the content word dictionary to determine the part of speech and accent type.

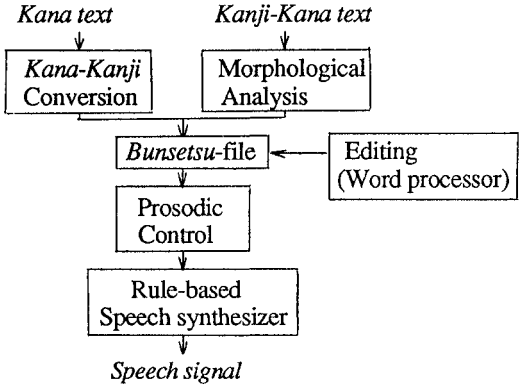


Fig . 1 Configuration of the system .

The assumption of properties of an undefined word expressed by Chinese characters is more difficult than that of other characters (*Hira-kana*, *Kata-kana*) because one Chinese character has several pronunciation.

The part of speech and accent type are determined by the properties of the single Chinese character, when a word is described by the Chinese characters. When a word is described by *Hira-Kana* or *Kata-Kana*, the accent type is determined by the number of morae. The properties of a single Chinese character and some examples of the definition of properties of undefined word are shown in Fig. 3.

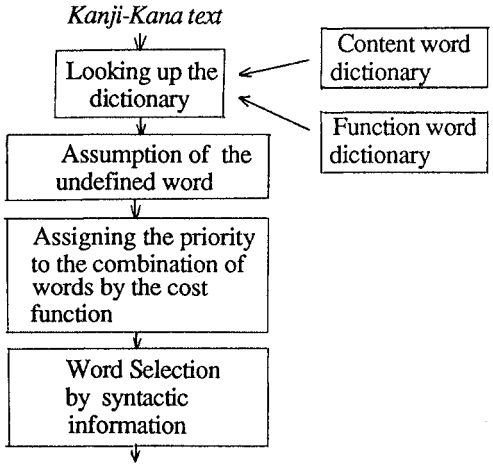


Fig. 2 Block diagram of the morphological analysis .

Symbol Proper noun First name

Noun Place name Family name

(1) Part of speech of a single Chinese character

「明」

part of speech	Proper noun	Family name	First name
Pronunciation	mei	ake	aki
Direction dependency of the combination with other character	no	backward	no
accent type	high	-	low

(2) Example of the properties of the single Chinese character

part of speech	expression	pronunciation with accentuation
Proper noun	明(mei) + 光	meikou
Family name	明(ake) + 井	akei
First name	明(aki) + 宏	aki'hiro
	善 + 明(aki)	yoshi'aki

(3) Example of the definition of the properties

Fig. 3 Examples of the determination of an undefined word.

3.3 Assigning the priority of the combination of words by cost function

As there are many word candidates, it is necessary to decrease the number of the candidates. In this process an order of priority is given to the combination of words. The cost function is used to give the priority to the combination of words. In general, the number of *Bunsetsu* has been widely used for the cost function. It is well known that the number of *Bunsetsu* is equivalent to the cost in the sentence if the cost of every content word is equal to 1, and that of every function word is equal to 0[3].

Though the number of *Bunsetsu* is a good approximation to the cost function for the Japanese sentence, the content word dictionary includes many suffixes are stored as one word. This means that the cost of each word is should not to be equal. The cost of the idiom and the prefix must be different.

Let us suppose that "日本人" (Japanese) is included in the dictionary and "アメリカ人" (American) isn't included in the dictionary. As the structure of both words is the same, the cost must be the same. In this system the cost of suffix "人" is equal to 0, consequently both words have a cost equal to 1. Examples of cost in compound words are shown below. The lower the cost of combination of words the higher the priority.

examples

日本人数学者
日本＋人数＋学者 total cost = 3
1 1 1 (cost of each word)
日本人＋数学＋者 total cost = 2
1 1 0
アメリカ人数学者
アメリカ＋人数＋学者 total cost =3
1 1 1
アメリカ＋人＋数学＋者 total cost = 2
1 0 1 0

The cost is used for not only the evaluation of the priority of *Bunsetsu* but also the determination of accent type in a compound word. The cost given to the compound word in the dictionary is greater than 1. The accent type of a compound word doesn't change when a compound word is combined with another word on the left hand side[4].

examples

	pronunciation with accent type
研究 (cost =1)	kenkyuu
研究所 (cost =2)	kenkyuujo :compound word
音声＋研究	onseike'nkyuu
音声＋研究所	onseikenkyuujo

3.4 Word selection by syntactic information

The correct words are selected from the candidates of the combination of words that gives the minimum cost in the sentence .

3.4.1 Syntactic and semantic features

Syntactic and semantic features are used in order to determine the word boundary and pronunciation. These features are given to approximately 10,000 words for a 55,000 word dictionary. The features used here are as follows: statistics of the word boundary, the dependency of direction of the combination in the compound word, lexical co-occurrence, case marker, and the frequency of use of the word. The features with some examples are shown in Fig. 4.

The statistics of the word boundary are given for the word which includes the same character as the prefix or suffix. This feature is used to determine the word boundary. The other features are used to select the correct pronunciation when the word boundary is the same. The dependency of the direction of the combination in the compound word is given by analyzing the word structure of the compound word.

3.4.2 Evaluation of the connectivity of the combination of words

The connectivity of the combination of words is evaluated. The system tests all features for each candidate. The connectivity is the sum of the values which are given to a candidate when the feature matches the condition.

a. Statistics of the word boundary
元首＋相 (rare)
元＋首相 (high frequency)
b. Syntactic information
b-1. Compound word or not
中 (naka) : uncombined
b-2. Direction of the combination
中 (chuu) + 程度 : combined with the right word
仕事 + 中 (chuu) : combined with the left word (verb)
日本 + 中 (juu) : combined with the left word (place)
b-3. Restriction of the case marker
行く : human + が, place + に
行なう : abstract + を
c. Co-occurrence
c-1. Reading of the Chinese character
船 (fune): Japanese pronunciation
船 (sen): Chinese pronunciation
c-2. Co-occurrence of the word category
Semantic features
time abstract
place,location concrete
measurable human
countable activity
Syntactic features
personal pronoun
abbreviation of the country name
number
unit
Pronunciation
number (Japanese style)
idiom
d. Frequency of use of the word

Fig. 4 Syntactic and semantic features.

When the cost of one sentence is equal to N, the sentence is divided into N sections. The evaluation is performed for every section.

The order of priority of each feature is as follows.

high order
(1) Co-occurrence of the word category
(2) Part of speech
| (2-1) Content word
| (2-2) Function word (Case marker)
| (3) Reading of the Chinese character
| (4) Direction of the combination
V (4-1) Compound word
(4-2) Single word
(5) Frequency of use of the word
low order

The word selection is performed from the beginning of the sentence to the end. If the words included in the Mth section are W_M ($M =< N$), the maximum value of connectivity with the M-1th section and Mth section is Va, and the pair of words that gives Va $W_{M,a}$ and $W_{M-1,a}$, that of Mth section and $M+1$th section is Vb $W_{M,b}$ and $W_{M+1,b}$ ($M >= 2$, if $M=N$, then $Cb = 0$).

if $Va >= Vb$
select $W_{M-1,a}$ $W_{M,a}$.
if $Vb < Va$
if
$W_{M,a} == W_{M,b}$
then
select $W_{M,a}$
else if
M-1th Word is not selected yet

example
米ソ中距離核全廃条約
section
1st 米ソ
2nd 中(中1:Chinese,中 2 :middle)
3rd 距離
4th 核
evaluation
2nd section
Pair a. 米ソ - 中1
$W1,a$ $W2,a$
Pair b. 中2 - 距離
$W2,b$ $W3,b$
$Va < Vb$
3rd section
Pair a. 中2 - 距離
$W2,a$ $W3,a$
Pair b. 距離 - 核
$W3,b$ $W4,b$
$Va > Vb$
selection of 1st to 3rd section
米ソ - 中2 - 距離
$W1,a$ $W2,a$ $W3,a$

Fig. 5 Examples of Bunsetsu selection .

then
select M-1th word that has the maximum value with M-2th section.

Examples of the selection are shown in Fig.5.

VI. RESULT OF THE EVALUATION TESTS

The evaluation tests for the sample article have been performed. The pronunciation, part of speech, and word boundary are checked. The score is calculated as a percentage of the number of correct *Bunsetsu* to the number of all *Bunsetsu* in the whole article. The result of the evaluation test of the sample article from the popular economical newspaper is shown in Table 1. The accuracy of the morphological analysis for the sample article is 99.0%.

Table 1 Result of the evaluation test.

Sentences 977 Bunsetsu	16,290
Correct Bunsetsu	16,125 (99.0%)
Errors (1.02%)	
Pronunciation	51 (0.31%)
Part of Speech	34 (0.21%)
Undefined word	38 (0.23%)
Boundary	42 (0.26%)

V. CONCLUSION

This paper describes the configuration of a linguistic processing module for Japanese text-to-speech system, the configuration of the morphological analysis module, the linguistic processing techniques employed, and the results of evaluation tests.

Syntactic and semantic features are used to decrease the error rate of the morphological analysis.The morphological analysis module performed with an accuracy of 99.0% when used to read a sample article from a popular Japanese newspaper.

ACKNOWLEDGEMENTS

The authors would like to express their appreciation to Director Ono of the KDD R&D Laboratories, to Deputy Director Urano, and to the members of Artificial Intelligence Group.

REFERENCES

[1] Y. Sagisaka and H. Sato, "Review of text-to-speech conversion technology," J. Acoust. Soc. Jpn.,Vol.41, pp.901-905, Dec. 1985 (in Japanese).
[2] S. Yamamoto, N. Higuchi and T. Shimizu, "Implementation of a Speech Synthesizer Based on Rule with Word Processing Functions," Trans. IEICE Jpn, Vol. J72-A, pp.273-284, Feb. 1989 (in Japanese).
[3] K. Shudo and K. Yoshimura, "The structure of Japanese Language and Its Mechanical Analysis," IPS Jpn, Vol 27, pp.947-954, Aug. 1986 (in Japanese).
[4] H. Sato, "Phonological Rules and Lexical Information for Speech Production," Spring National Convention Record, SA-10, pp.359-360, Apr. 1988 (in Japanese).