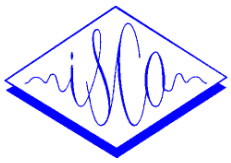

First International Conference on
Spoken Language Processing (ICSLP 90)
Kobe, Japan
November 18-22, 1990

A SPEECH RECOGNITION RESEARCH ENVIRONMENT BASED ON LARGE-SCALE WORD AND CONCEPT DICTIONARIES

Hiroyuki Chimoto Hideaki Shinchi** Hideki Hashimoto** Shinya Amano**

* Japan Electronic Dictionary Research Institute, Ltd.
Fifth Research Laboratory Research Center
** Toshiba Software Engineering Corporation
c/o Toshiba Corporation Research and Development Center
1 Komukai Toshiba-cho, Saiwai-ku, Kawasaki 210, Japan

Abstract

This paper describes a speech recognition research environment based on large-scale Japanese word and concept dictionaries. The dictionaries are currently under development at Japan Electronic Dictionary Research Institute (EDR). The research environment includes user-interactive capabilities, which enable efficient upgrade and revision of the dictionaries; and it has high-speed hardware, which supports real-time processing of speech input data. Within this environment, a combination of the two dictionaries will be applied to deal with the inherent ambiguities of speech recognition.

I Introduction

Several language processing systems, employing dictionaries and knowledge bases, have been developed for both machine translation and speech recognition applications [1]. These systems work well in restricted environments; however, they are task-oriented and use complex knowledge representations. This creates obstacles when it is necessary to integrate updated knowledge into the system.

In order to avoid these obstacles, the Japan Electronic Dictionary Project was initiated at EDR in 1986. The goal of the project is to develop a large-scale electronic dictionary for natural language processing. Both word dictionaries and concept dictionaries are included in the project.

In addition, the project promotes the development of evaluation systems. These are being used to evaluate the electronic dictionary in high-level natural language processing applications, including machine translation and speech recognition/generation. The development of evaluation systems are crucial to the creation of large-scale knowledge bases; furthermore evaluation of the dictionaries is very difficult. For example, when an incorrect information description is discovered and revised, it is necessary to determine if this description influences additional descriptions. The descriptions that are affected by the revision must be changed accordingly.

In general, maintenance of a large-scale knowledge base is very difficult. The CYC project is being done at MCC to build a large-scale knowledge base. In order to maintain the knowledge base, CYC consists of not only the knowledge base itself but also an environment which supports interface editing and browsing tools [2]. This environment is a crucial element of CYC because it enables maintenance and upgrade of the knowledge base. In the same light, the authors have implemented a special window configuration within the speech recognition environment for the same purpose.

This paper describes the configuration of the EDR electronic dictionary and our speech recognition research environment.

II EDR Electronic dictionary

The EDR electronic dictionary [3, 4] is composed of word dictionaries, concept dictionaries, co-occurrence dictionaries, and bilingual dictionaries. There is a total of ten dictionaries as shown in Figure 1. The word dictionary [5] consists of the general vocabulary dictionary and the technical terminology dictionary based on lexical differences.

The general vocabulary dictionary is composed of two general vocabularies, English and Japanese. Each word dictionary consists of 200,000 words. The technical terminology dictionary is used for information processing. It also has both Japanese and English vocabularies, each containing 100,000 words. These word dictionaries include grammatical information used as clues to determine the syntactic structure and description represented by words. Therefore, these word dictionaries are used for morphological and syntactic processing.

The concept dictionary [6] consists of knowledge about the 400,000 concepts defined by the word dictionary. It has two types of dictionaries, concept classification and concept description. These concept dictionaries are used for semantic processing. The concept dictionary is crucial for enabling computer understanding of concepts described in the word dictionary. Just as humans use a reserve of knowledge to understand the concepts that comprise a sentence, the concept dictionary provides computers with such knowledge.

EDR is also developing, the co-occurrence dictionary and the bilingual dictionary. Although the Japanese word dictionary and concept dictionary are currently being evaluated in speech recognition applications, the co-occurrence dictionary and the bilingual dictionary are not.

III Evaluation System

3.1 System configuration

To evaluate the effectiveness of the Japanese word and

concept dictionaries, a speech recognition research environment has been constructed as shown in Figure 2. The environment includes two general purpose workstations. Workstation#1 has two special-purpose, high-speed processor boards [7]. The function of these boards is to enable real-time processing at the signal level. The environment supports multiple windows for observing intermediate processing results. These windows can display each recognition candidate and its similarity value at any stage during the recognition process, and they allow on-screen dictionary and grammar modification as shown in Figure 2.

3.2 Processing flow

This evaluation system supports both low-level and high-level functions.

At the low level, real-time signal analysis and recognition are performed. Since the language used for evaluation is Japanese, the basic unit of input speech is a mora, or syllable unit. Analysis and recognition of the input syllable is carried out by the processor boards. Syllable candidates from the recognition result are sent to Workstation#2. The recognition system employs the Multiple Similarity Method [8].

At the high level, morphological and syntactic analysis is performed. Using a beam-search method, Workstation#2 executes morphological analysis on the syllable candidates. After all of the syllable candidates have been individually analyzed, Workstation#2 then conducts syntactic analysis on phrase candidates with a Generalized L-R parser.

The budge processing is performed by the software. The evaluation sentence is entered into the system from the keyboard, or it is read from the data-base file. Workstation#2 generates syllable lattices based on a confusion-matrix and a distribution of similarity values.

The evaluation sentences are divided into two types, literary style and colloquial style. The literary style sentences are extracted from newspaper articles, and the colloquial style sentences are questions corresponding to the articles.

IV Evaluation environment

4.1 Windows

A special feature of this system is the ability to observe results at each stage of analysis using the multiple windows of Workstation#2. Each window corresponds to a distinct analysis stage, and the results are presented at the end of each stage.

The evaluation window consists of a syllable candidate window, word candidate window, phrase candidate window,

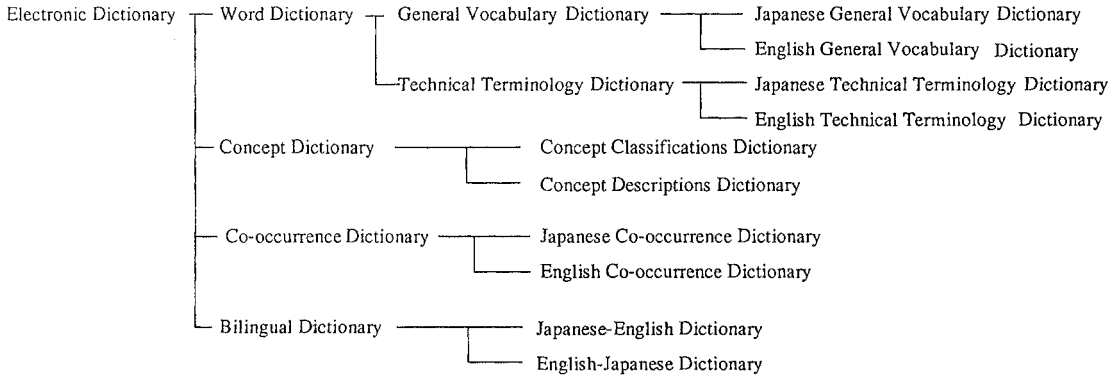

Figure 1 Structure of the EDR electronic dictionaries

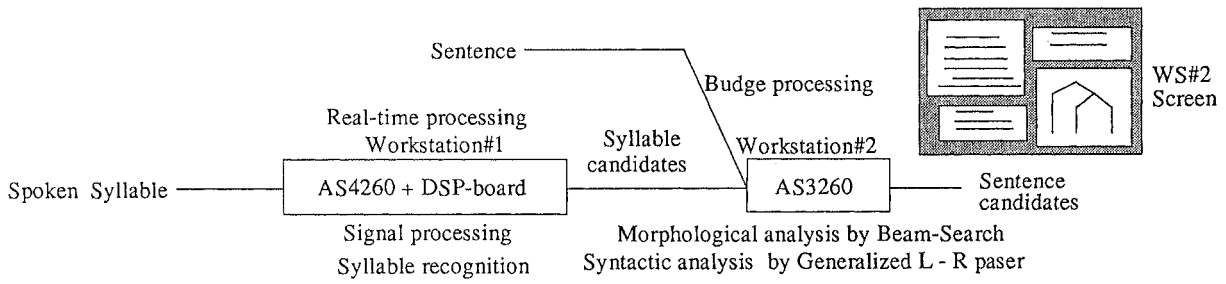

Figure 2 Block diagram of the evaluation System

parse tree window and special-purpose work window as shown in Figure 3.

4.1.1 Syllable candidate window

This window displays syllable candidates and their similarity values for each input syllable. The main function of this window is to confirm the recognition result. All of syllable recognition results are used as candidates. If all of syllable candidates are not used, it is possible that the correct answer will not be found. These syllable candidates are displayed in phrase units.

4.1.2 Word candidate window

This window displays each of the word candidates as well as its part of speech information and candidate score. The word candidates are obtained by extra notation of the dictionary. An objective of this window is to examine part of speech information and the extra notation of the word dictionary.

4.1.3 Phrase candidate windows

Similar to the word candidate window, this window displays phrase candidates which are obtained by the beam-search method using part of speech information and left-side and right-side adjacency attributes. The phrase candidate window also shows left-side and right-side adjacency attributes.

4.1.4 Parse tree window

This window displays parse tree candidates and part of speech information of words in the phrase. The trees are obtained through syntactic analysis by a Generalized L-R parser. The main purpose of this window is to examine the developing grammar.

4.1.5 Work window

This window is a special-purpose window. When one of the processes is unable to successfully analyze the input data, the corresponding window presents a mark of failure (X). All windows corresponding to subsequent processes also display the mark. Therefore, the user can easily find the process in which the failure occurred. The user can then select a grammar or dictionary window in order to modify the grammar or dictionary.

4.2 Example

If a failure is found in the word dictionary using the above environment, revision of the dictionary is performed as follows.

At first, the syllable candidate window presents a mark of failure (X) for morphological analysis as shown in Figure 4-a. After that, the word dictionary or grammar window is selected with the mouse. Using the headword of input, wrong information is searched by mouse (Figure 4-b). Each item is checked, and the new information is written (Figure 4-c). In the next step, the new information is tested on the same sentence. If the test is a success, the new information is entered into the history file. The relationship defined by this new information must not contradict existing relationships among other items or information. Therefore, the new information is tested throughout the database by budge processing.

After that, if the new information succeeds in the budge

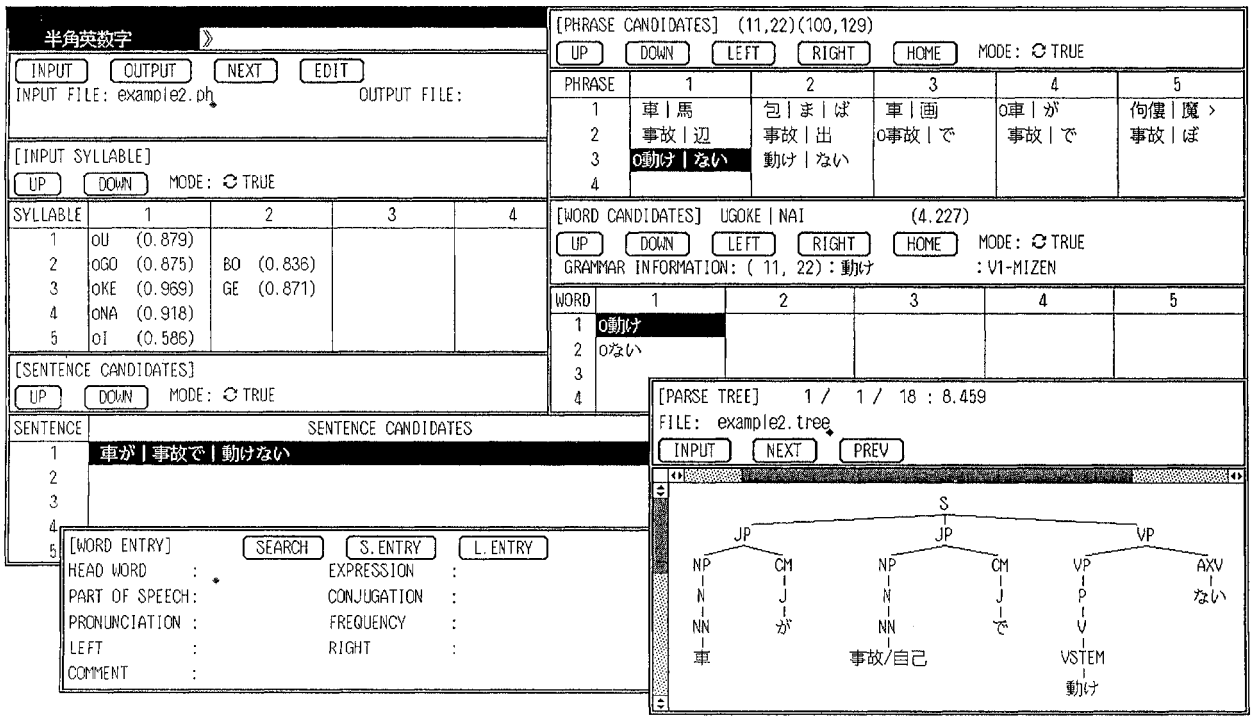

Figure 3 Evaluation screen

processing, it is written in the dictionary. Because this information may have harmful effects in another test, the user must note user-name, date, and reason for the rewrite in the history file. If the user does not write these comments, the system displays a warning message, and the new information is not entered into the dictionary or grammar file. If the user writes the comments, the information concerning dictionary modification is recorded in the history file; therefore, the user can check this information at any time (Figure 4-d).

V Conclusion

For the purpose of supporting the improvement of EDR dictionaries and of speech recognition performance, a speech recognition research environment has been developed. Because of the user interactive features using multiple windows and high-speed processor boards, researchers can simultaneously observe and maintain multiple levels of information for efficient evaluation and upgrade of the system including dictionaries and grammar.

References

[1] V. R. Leser, R. D. Fennell, L. D. Erman and D. R. Reddy, "Organization of the HearsayII Speech Understanding System," IEEE Trans. ASSP-23, 1, pp. 11-24, 1975.

[2] D.B. Lenat and R.V. Guha, "Building Large Knowledge-Based Systems - Representation and Inference in the CYC Project," Addison-Wesley, 1989.

[3] EDR, "An Overview of the EDR Electronic Dictionaries," EDR Technical Report, TR-024, 1990.

[4] T. Yokoi, H. Uchida, S. Amano, and M. Kiyono, "Research and Development of Large-Scale Electronic Dictionaries," The First Australia - Japan Joint Symposium on Natural Language Processing, Australia, 1989.

[5] EDR, "Word dictionary (Ver. 2)," EDR Technical Report, TR-008, 1988.

[6] EDR, "Concept Dictionary," EDR Technical Report, TR-027, 1990.

[7] H. Chimoto and H. Tsuboi, "Development of Fast DSP board for Speech Recognition Research," Record of 1989 Fall Joint convention of Acoust. Soc. Japan, 3-4-10, 1989. (In Japanese)

[8] Y.Takebayashi, H.Shinoda, T.Nitta, S.Hirai and S.Watanabe, "Telephone Speech Recognition Using a Hybrid Method," Proc. Seventh IJCPR, pp. 1232-1235, 1984.

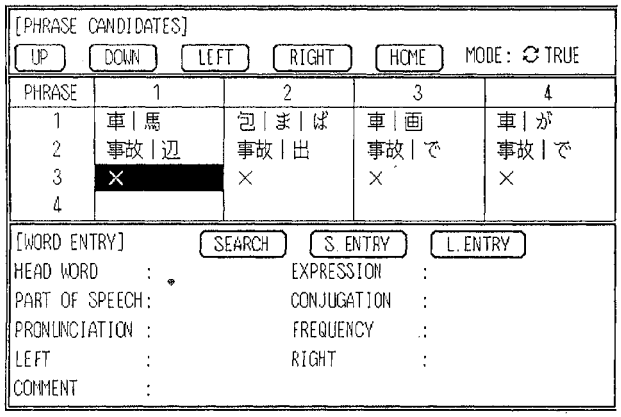

Figure 4 - a Work window
Selecting dictionary information

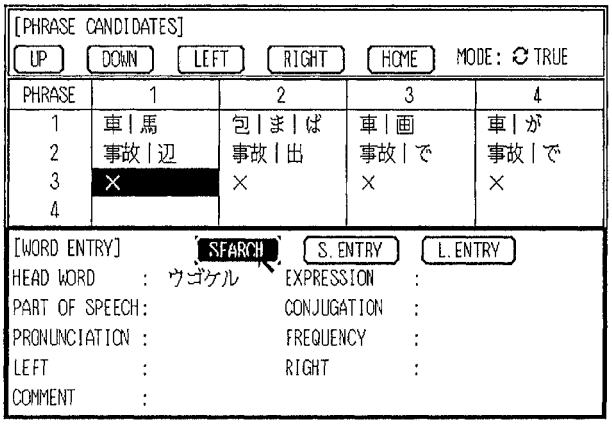

Figure 4 - b Work window
Selecting error item

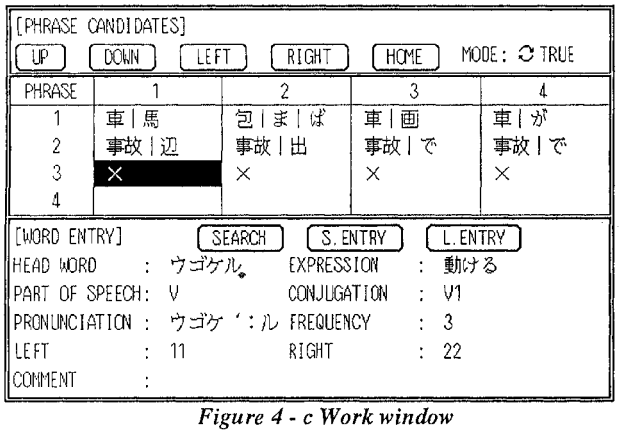

Figure 4 - c Work window
Rewiting a new information

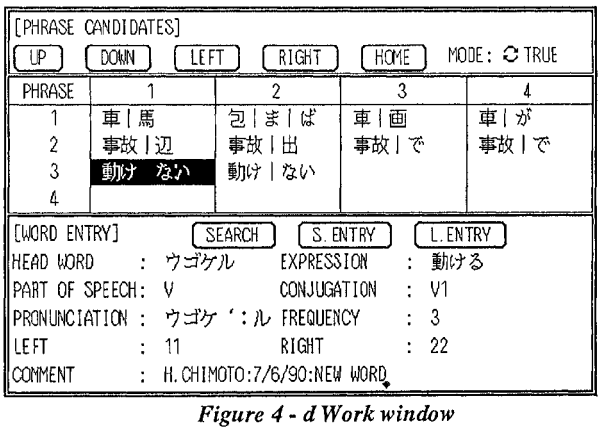

Figure 4 - d Work window
Reading a history file