

# A Text Input Front-end Processor as an Information Access Platform

**Shinichi DOI, Shin-ichiro KAMEI and Kiyoshi YAMABANA**

C&C Media Research Laboratories, NEC Corporation

4-1-1, Miyazaki, Miyamae-ku, Kawasaki, KANAGAWA 216-8555 JAPAN

s-doi@ccm.cl.nec.co.jp, kamei@ccm.cl.nec.co.jp, yamabana@ccm.cl.nec.co.jp

## Abstract

This paper presents a practical foreign language writing support tool which makes it much easier to utilize dictionary and example sentence resources. Like a Kana-Kanji conversion front-end processor used to input Japanese language text, this tool is also implemented as a front-end processor and can be combined with a wide variety of applications. A morphological analyzer automatically extracts key words from text as it is being input into the tool, and these words are used to locate information relevant to the input text. This information is then automatically displayed to the user. With this tool, users can concentrate better on their writing because much less interruption of their work is required for the consulting of dictionaries or for the retrieval of reference sentences. Retrieval and display may be conducted in any of three ways: 1) relevant information is retrieved and displayed automatically; 2) information is retrieved automatically but displayed only on user command; 3) information is both retrieved and displayed only on user command. The extent to which the retrieval and display of information proceeds automatically depends on the type of information being referenced; this element of the design adds to system efficiency. Further, by combining this tool with a stepped-level interactive machine translation function, we have created a PC support tool to help Japanese people write in English.

## 1. Introduction

When creating text using word processing software on a personal computer, it is common to refer to books or documents relevant to the text, including various kinds of dictionaries and reference works. The tools used for accessing relevant information, such as CD-ROM dictionaries, text databases, and text retrieval software, however, often require user actions that may seriously interrupt the writing process itself. These may include executing retrieval software, inputting key words, or copying retrieved information into texts.

The foreign language writing support tool we propose here automatically access information relevant to input texts. Like a Kana-Kanji conversion front-end processor used to input Japanese language text, this tool is also implemented as a front-end processor (FEP) and can be combined with a wide variety of applications. The extent to which the retrieval and display of information proceeds automatically depends on the type of information being referenced; this element of the design adds to system efficiency.

In Section 2, we consider the requirements for efficient writing support tools and discuss the characteristics of our front-end processor and its automatic information access function. In Section 3, we introduce our English writing support tool, which has been developed to help Japanese people write in English on a PC. This tool combines a front-end processor with the stepped-level interactive machine translation method we first proposed in Yamabana (1997). In Section 4, we describe the automatic information access function of the English writing support tool.

## 2. FEP-type Information Access Platform

### 2.1. Text input front-end processor with information access functions

To allow users to concentrate better on their work, writing support tools with reference information access functions should:

- 1) provide for automatic access of reference information, i.e. access without explicit user commands,
- 2) enable users to utilize retrieved information with simple operations, and
- 3) be compatible with a wide variety of word processing applications.

In developing our FEP-type support tool, we started with the text retrieval application proposed in Muraki (1997), which provides a morphological analyzer that automatically analyzes users' input and extracts key words to retrieve relevant text from a database. This application fulfills the first of the requirement listed above. We converted such a morphological analyzer into an FEP for use in our tool, which is placed between the keyboard and an application. When a user inputs texts into this tool, the morphological analyzer identifies each word and extracts key words automatically before the text is entered into the application. The key words are used to retrieve information relevant to the input texts. This information is displayed for easy editing and utilization. Because all of this can be achieved with standard hooks and the IME API of the Microsoft Windows 95 operating system, this tool can be combined with any Windows-compatible text-input application. In addition, it can be combined with any other front-end processor, including Kana-Kanji conversion FEPs, through the use of a technique we have recently developed. Figure 1 shows the tool architecture.

### 2.2. Controlling the extent of the automation of information retrieval and display

The automatic retrieval and display function introduced in the previous subsection allows users to concentrate better on their writing

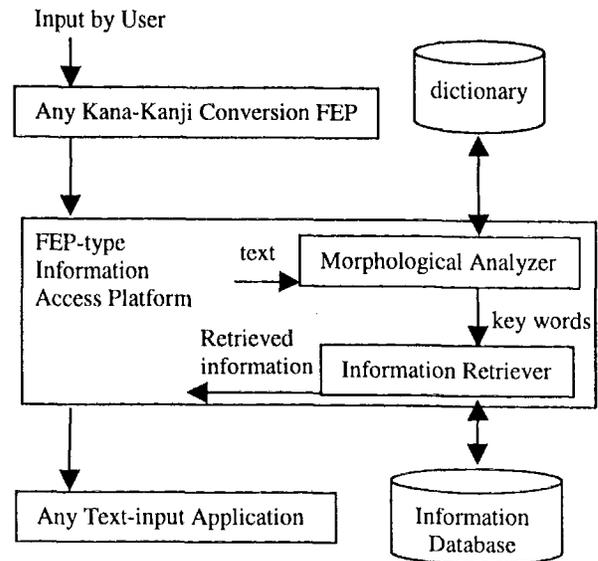


Figure 1 Architecture of the FEP-type Information Access Platform

because much less interruption of their work is required for the consulting of dictionaries or for the retrieval of reference sentences. This function, however, might prevent users from concentrating on their writing if all the retrieved information were displayed in a new window, especially when the quantity of the retrieved information were large and the majority of it were not relevant from the users' point of view.

To compensate for this disadvantage, we divided the information access function into three steps: 1) extracting key words from the input text, 2) using the key words to retrieve reference information, and 3) displaying the retrieved information, and we developed a function to control whether the each step is executed automatically or manually. We prepare three methods for retrieval and display as follows.

- A) Relevant information is retrieved and displayed automatically, without user command.
- B) Information is retrieved automatically but displayed only on user command. After automatic retrieval, only the quantity of information is displayed, and users can decide whether to display it.
- C) Information is both retrieved and displayed only on user command. Even in this case, because key words are automatically

extracted before retrieval, our tool requires much less user action than other information accessing tools.

The extent to which the retrieval and display of information proceeds automatically depends on the type of information being referenced; this element of the design adds to system efficiency.

### 3. English Writing Support Tool “Eibun Meibun Meikingu”

By combining the FEP-type information access platform with the stepped-level interactive machine translation method we proposed in Yamabana (1997), we have developed an English writing support tool to help Japanese people write in English on a PC. This tool, named “Eibun Meibun Meikingu”<sup>1</sup>, consists of the following three components:

- 1) an English writing FEP, “Eisaku Pen”<sup>2</sup>, which converts Japanese into English,
- 2) a CD-ROM dictionary consulting tool, “Shoseki Renzu”<sup>3</sup>, and
- 3) a Japanese-to-English bilingual example sentence database, “Reibun Baina”<sup>4</sup>.

Figure 2 shows the architecture of “Eibun Meibun Meikingu”. This tool is now available as a software package.

#### 3.1. English writing FEP “Eisaku Pen”

“Eisaku Pen” has an interactive interface similar to Kana-Kanji conversion FEPs, and initially replaces most of the Japanese vocabulary items with English equivalents but maintains Japanese grammatical constructions. When a user inputs Japanese text, a conversion window of “Eisaku Pen” is automatically popped-up and English equivalents are displayed in the order of original Japanese words. Figure 3 illustrates how text is

<sup>1</sup> The Japanese terms *Eibun*, *Meibun* and *Meikingu* mean, respectively, ‘English writing’, ‘beautiful writing’ and ‘making’.

<sup>2</sup> The Japanese terms *Eisaku* and *Pen* mean, respectively, ‘Creating English’ and ‘a pen’.

<sup>3</sup> The Japanese terms *Shoseki* and *Renzu* mean, respectively, ‘written materials’ and ‘a lens’.

<sup>4</sup> The Japanese terms *Reibun* and *Bainda* mean, respectively, ‘example sentences’ and ‘a binder’.

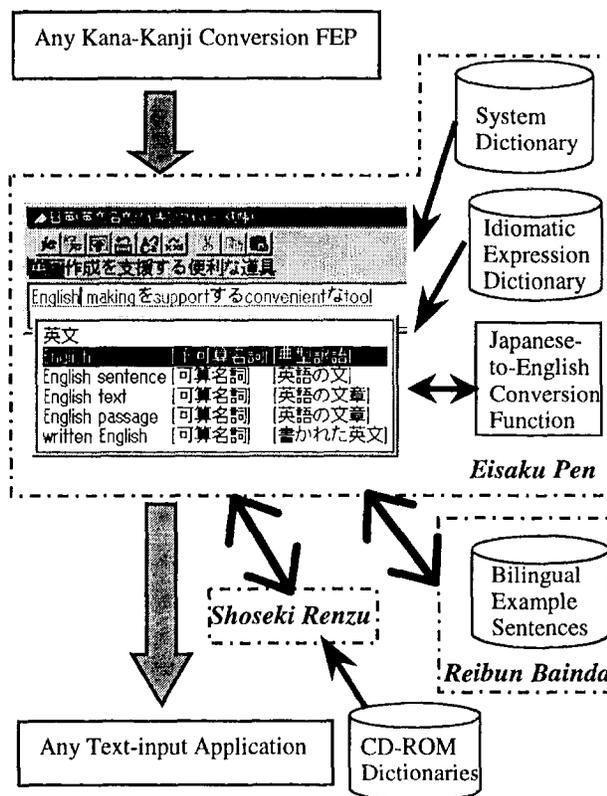


Figure 2 Architecture of the English Writing Support Tool “Eibun Meibun Meikingu”

displayed. When a user inputs Japanese sentence “purezento wo arigato”, where each word means ‘present’, objective marker and ‘thank you’ respectively, “purezento” and “arigato” are replaced with their English equivalents ‘present’ and ‘thank you’ and displayed automatically in the conversion window shown in the center of the



Figure 3 Illustration of “Eisaku Pen”

figure. The window below is an alternatives window to display all the possible equivalents for “*arigato*”, by selecting from which, users can easily change equivalents. In this alternatives window, “*Eisaku Pen*” provides part-of-speech of each alternative equivalents and supplementary information indicating the difference between their meanings or usage in order to make users’ equivalent selection easier.

After confirming the equivalents of input words, users can execute the Japanese-to-English conversion function, which transforms Japanese grammatical constructions into those of English and the whole sentence is converted to an English sentence: ‘Thank you for a present.’ by automatic word reordering and article insertion. This syntactic transformation proceeds step by step, in a bottom-up manner, combining smaller translation components into larger ones. Such a ‘dictionary-based interactive translation’ approach allows users to refine dictionary suggestions at different steps of the process. Finally, users can also easily change articles to obtain the result sentence: ‘Thank you for the present.’

The system dictionary of “*Eisaku Pen*” contains about 100,000 Japanese vocabulary entries and 15,000 idiomatic expressions. Since there was no source available to build an idiom dictionary of this size, we collected them manually, from scratch, following a method described in Tamura (1997).

### 3.2. CD-ROM dictionary consulting tool “*Shoseki Renzu*”

While using “*Eisaku Pen*”, if users want to obtain more information on words or equivalents, “*Shoseki Renzu*” provides a function to consult CD-ROM dictionaries.

For example, when users execute the CD-ROM dictionary consulting function of “*Shoseki Renzu*” at the situation of the Figure 3, the currently selected alternative ‘thank you’ is regarded as a key word for dictionary consulting and the contents of the dictionaries for ‘thank you’ is displayed. If users double-click on another word in a conversion window or an alternatives window including the original Japanese word shown at the top of the window,

the word is regarded as a key word for dictionary consulting.

### 3.3. Bilingual example sentence database “*Reibun Baida*”

“*Eibun Meibun Meikingu*” also provides a function to retrieve and utilize bilingual example sentences. Example sentences relevant to the texts input by users are retrieved from the database of “*Reibun Baida*” containing 3,000 of Japanese-to-English bilingual sentence pairs for letter writing. Figure 4 illustrates the Japanese-to-English sentence pairs retrieved when a user executes “*Reibun Baida*” at the situation of the Figure 3. Here, the currently selected original Japanese word “*arigato*” is regarded as a key word for retrieving and the example sentences which are assigned a key word “*arigato*” beforehand or include strings of “*arigato*” in the Japanese sentence are retrieved from the bilingual example sentence database of “*Reibun Baida*” and displayed in the window as illustrated in Figure 4. Japanese sentences are shown in the first column and translated English sentences are shown in the second one. The third one is for supplementary information indicating the difference between meanings or usage of the sentences. Users can easily send these sentences to text-input applications by drag-and-drop operation using a mouse. In addition, by using “*Eisaku Pen*”, users easily edit a Japanese word and its English equivalents in example sentences synchronously.

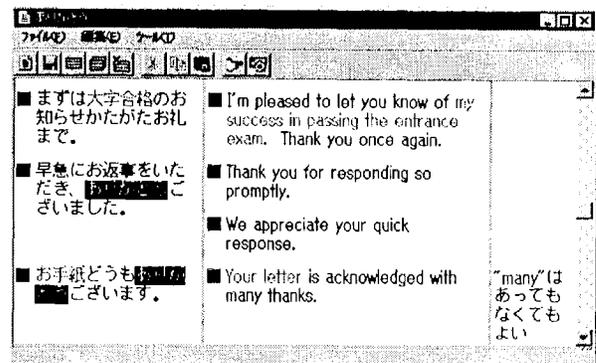


Figure 4 Illustration of bilingual sentences retrieved by “*Reibun Baida*”

#### 4. Information Access Function of English Writing Support Tool

Our tool currently accesses three types of information: 1) information, included in the system dictionary, regarding grammatical forms and idiomatic expressions; 2) straight CD-ROM dictionary information; and 3) Japanese-to-English example sentences in the database. The extent to which the retrieval and display of information proceeds automatically depends on the type of information being referenced; information of type 1) is retrieved and displayed automatically, that of type 2) is both retrieved and displayed manually, and that of type 3) is retrieved automatically but displayed manually.

In the first case of translation equivalents and grammatical information retrieval, “*Eisaku Pen*” automatically retrieves and displays English words equivalent to the input Japanese texts without explicit user command because users always utilize the English equivalents in English writing.

In the second case of CD-ROM dictionary consulting, “*Shoseki Renzu*” retrieves and displays contents of CD-ROM dictionaries on user command because this dictionary consulting function needs to be executed only when users require additional information. Our tool requires much less user action than other dictionary consulting tools because key words are automatically extracted before user command for retrieval and users don’t always need to input key words.

In the third case of bilingual sentence retrieval, “*Reibun Baimda*” retrieves sentences automatically but displays only on user command. Because “*Reibun Baimda*” contains the example sentences in itself, relevant sentences are retrieved at high speed and the retrieval function doesn’t interrupt users’ writing process. Retrieved sentences, however, might include the ones not relevant to the input text from users’ point of view, because similarity between sentences is judged with a simple method using key words. Therefore, the writing process might be interrupted if retrieved sentences were displayed automatically. To avoid this problem, the color of the icon of “*Reibun Baimda*” is changed after automatic retrieval, depending on

the existence of relevant sentences, and users can decide whether to display the retrieved sentences.

#### 5. Conclusion

We present a practical foreign language writing support tool which makes it much easier to utilize dictionary and example sentence resources. This tool is implemented as a front-end processor and can be combined with a wide variety of applications. The extent to which the retrieval and display of information proceeds automatically depends on the type of information being referenced; this element of the design adds to system efficiency. We also describe our English writing support tool with a stepped-level interactive machine translation function, by which users can write English by accessing essential information resources including bilingual dictionaries and example sentences.

Our tool is implemented as an English writing support tool, now under expansion to a general writing support tool. Another further work is enlarging resources our tool can access. We are also developing an example-based translation function which utilizes example sentences in “*Reibun Baimda*” for Japanese-to-English conversion function of “*Eisaku Pen*” and an automatic example sentence acquisition function which acquires users’ input texts and their translation and adds them to “*Reibun Baimda*” automatically.

#### References

- Muraki K., et al. (1997) *Information Sharing Accelerated by Work History Based Contribution Management, Leads to Knowhow Sharing*. In “Design of Computing Systems: Cognitive Considerations”, Salvendy G., et al. ed., Elsevier Science B.V., Amsterdam, pp. 81-84.
- Tamura S., et al. (1997) *An Efficient Way to Build a Bilingual Idiomatic Lexicon with Wide Coverage for Newspaper Translation*. NLPRS’97, Phuket, Thailand, pp. 479-484.
- Yamabana K. et al. (1997) *An Interactive Translation Support Facility for Non-Professional Users*. ANLP-97, Washington, pp. 324-331.