

TWP: How to Assist English Production on Japanese Word Processor

Kazunori MURAKI, Susumu AKAMINE†, Kenji SATOH and Sinichi ANDO

Human Language Research Lab., Information Technology Research Labs., NEC Corp.

†Interpreting Telecommunications Research Labs., ATR

Miyazaki 4-1-1, Miyamae-ku, Kawasaki 216 JAPAN *e-mail: k-muraki@hum.cl.nec.co.jp*

1 A new type translation support is available.

We propose the third type of translation support system in this paper, which enables users to enjoy simultaneously both large man-hour reduction provided by the conventional automatic machine translation system, and contentment with the functions brought by the dictionary reference and grammatical consultation. Our new system possesses the following features[4, 5, 6]:

1. it automatically generates the corresponding English expressions, simultaneously with the execution of inputting each Japanese word, phrase or sentence.
2. It assists word to word, or phrase to phrase translation, which allows users to control translation process.
3. It minimizes the number of key strokes for English translation with an aid of fully automatic phrase and simple sentence translation capability.
4. It guarantees that users can be content with an interface by which they can monitor and cancel each process of recognition and conversion of every input words, phrases, and sentences.
5. It greatly assists Japanese native speakers in preparing English document.

Let us call it "Translation Word Processor(TWP hereafter), realizing a very comfortable working platform equipped with a simultaneous automatic Japanese-English conversion facility that contributes to a drastic reduction of key stroke and mouse clicks.

The first type of translation support system or machine translation(automatic translation) system analyzes and generate sentences one by one. Even now when more than ten sorts of machine translation systems are available in the market, all we are forced to do while waiting for the translation to be outputted is to pray that the result may not be like the pieces of wreckage[1]. Even if it is not wrecked, we need to recognize damage and spend extra cost fixing it.

The second type is the bilingual dictionary consultation facility. This type of assistance systems are already in use on various kinds of word processors[10],

but does not achieve satisfactory efficiency in translation job. It demands users to set up the dictionary consultation function, to choose suitable expression out of the displayed words, phrases, and sample sentences, and to arrange them by key and mouse operations to produce English sentences. Even if dictionary reference is quickly and cheaply assisted, most part of English composition work is left in user's hands.

The third type ,TWP, is designed to visualize the automatic Japanese-English conversion(J/E conversion hereafter) process in parallel to inputting Japanese text, and to consequencely achieve the best cost-efficient performance in a frame work of man and computer interaction. In the TWP, various supporting functions are provided according to user's English writing capability:

1. J/E conversion of words, phrases and simple sentences,
 2. pattern based J/E conversion of complex sentences,
 3. Japanese-English translation DB retrieval,
 4. learning function of word and phrase conversion knowledge, and complex pattern conversion knowledge, which helps user to compose English greatly,
 5. automatic ambiguity resolution in conversion and retrieval designed to reduce the number of operations necessary for utilizing those assistance functions.
- For the purpose to achieve user's contentment in TWP operation,
6. flexible "cancel, "undo and "option selection interface are employed.
- The accuracy of ambiguity resolution,
7. rises up greatly by continual use of the learning function 4.

A Japanese structural editor with functions[7] stated above, can recognize a logical unit of a word, a phrase and a sentence in sequences of Japanese symbols. The Japanese-English conversion can start automatically at every moment a Japanese word, a phrase or a simple sentence (called logical units) is recognized. When more than two simple sentences(clauses) are extracted,

they are transferred to complex sentence conversion based on user definable complex sentence J/E translation patterns.

TWP can make user reach a target English by their monitoring and controlling Japanese-English conversion process, so that they can feel content with each step of every interactive assistances or automatic processes. It can also save key strokes for arranging word order as it generates English expressions in a correct word order by conversion capability of a phrase, a simple sentence and a complex sentence, besides word level dictionary consultation facility. TWP also has proposed a protocol to eliminate redundant user key strokes for selecting menu or commanding process by default reasoning to guess user's next stroke. This means that no other key strokes are sometimes necessary than those for inputting Japanese sentences, as translation is over at the same time of that a period of a Japanese sentence is keyed-in on a screen.

Whenever any error of conversion is found while inputting Japanese, any step during the conversion can be cancelled and corrected by adopting the second alternative automatically based on key stroke history learning mechanism. A cheap cancel and undo protocol is the key issue to make the best use of each default reasoning of TWP system, because strokes to select correct one after cancels and undos are the sources for tuning up default reasoning by learning mechanism.

2 What is considered of for comfortable usability?

TWP brings about both users' contentment with functions that machine translation has failed to realize, and large cost reduction for sentence generation in target language that simple dictionary consultation facility has not fulfilled. No other system has ever been presented to satisfy two of them at the same time. Machine translation system takes no human cost if complete translation is acquired. If any errors occur, however, manual re-translation or post-editing charges vast amount of expense. Post-editors have reported so many time high stress by repetitious or utterly unexpected errors machine translation caused. Moreover, even when they use Japanese-English automatic translation system, they are required to comprehend both languages in reality. Experience of using dictionary consultation facilities and grammar checkers tells us that such assistance as bilingual dictionary reference, sample sentence reference and grammatical check does not contribute so much to efficiency of translation and that the key operation cost taken to call those functions, e.g. key strokes, is essentially no less than that in case of consulting paper dictionaries.

TWP performs automatic recognition of a word, a phrase, and a simple sentence, and immediately after that, successive conversion based on bilingual word dictionary, and simple phrase and simple sentence

translation function is executed. These recognition and conversion processes are supervised by the user in the course of inputting Japanese text. TWP users reserve the means of cancelling and resolving the errors that the recognition or conversion created. TWP makes users feel content with processes on a screen by confirming every step of translation with their eyes. TWP supports so much easy use of dictionary consultation facility in such a manner that either English and Japanese word on a cursor position is interpreted as a key word for looking up word and bilingual dictionaries, and is transferred any time automatically to consultation facility to retrieve and display the dictionary data on a pull-down window. Users need not to take a special dictionary look up procedure.

It assists longer sentence generation in English word order by J/E conversion using complex sentence translation patterns defined by users[5], as well as using a phrase and a simple sentence translation facility.

In the following chapter, we explain the overview of TWP; English producing support system at first then talk about Japanese-to-English interactive conversion framework for comfortable English composition in particular.

3 Which Resources support English Composition ?

TWP for English composition support is constructed from well understood and transparent resources and mechanisms under an extended, what is called, Kana-Kanji conversion user interface. The noble feature of the TWP introduced hereafter, is of great transparency in every process of support mechanism. The traceability of process and the controllability of the TWP based on default reasoning mechanism in recognition and conversion of logical units in Japanese input, and cheap cancellability of wrong steps selected by TWP or a user are a noble part of invention which supports a great cost reduction of English production on Japanese word processors with full of contentment.

The characteristics of TWP is summarized as follows.

1. Japanese to English conversion interface simulating Kana-Kanji Conversion helps an user to learn so easily[1, 3, 4].
2. User controllable complex sentence translation pattern based conversion achieves translation transparency even in long sentence J/E conversion[5, 8, 9].
3. Automatic learning feature based on user key stroke history covers learnings for translation disambiguations of words, verb phrases, complex sentential conversion patterns, etc[6].
4. It is fully equipped with consultation facilities of translation examples, word dictionaries, bilingual

dictionaries and so on.

The figure 1 depicts the configuration of TWP prototype on a structural editor¹, with the PIVOT translation system as a back end conversion processor.

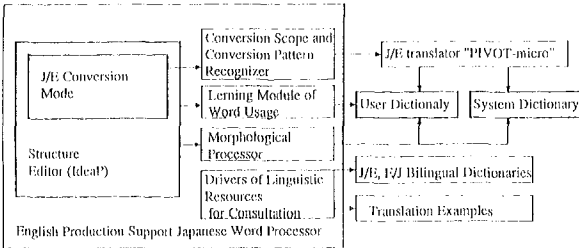


Figure 1. The configuration of TWP

The conversion of Japanese into English expression starts here with determination of scope to convert into English. The sequences of key strokes are transferred to morphological processing function for detecting logical units like words, phrases, clauses, etc. Any J/E conversion of each unit is carried out easily with MT subset of PIVOT J/E translation system as a back end translator.

The structure editor (IdeaP) on a robust Japanese morphological processing function detects Japanese logical units and keeps every strokes and every processes, during editing, conversion (translation), dictionary look up etc., in order for user to back to any point from unexpected result in English production by cancelling each preceding strokes and processes.

The cancelling protocol between man and computer plays a great role of providing users with the most comfortable operationability, which is backed up by the high transparency of system process.

The transparency introduced here is composed of two aspects; 1) hiding primitive mechanical steps from a screen and showing every and only logical unit level steps of edit and conversion on a screen, 2) letting even phrase, clause J/E conversion to look like Kana-Kanji conversion to casual users. The similarity of user interface protocol of the one we propose here and the widely acknowledged Kana-Kanji conversion interface lets users to believe in controllability of J/E conversion. The controllability of J/E conversion on Japanese word processors is easily obtainable through the experiences of Kana-Kanji conversion built in Japanese word processors.

¹The structure editor called "IdeaP"[7] can maintains the logical units like words, phrases, clauses and permits users to edit those units so easy while ordinal Japanese word processors cannot handle those logical units other than characters.

4 TWP operation and process

4.1 What meets casual user's requirement?

There is a big difference between J/E conversion and Kana-Kanji conversion. The former requires structural transformation like word order transformation, structural reduction or expansion, and so on, while the latter only handles a Kana sequence to produce a kanji sequence without any modification of the word order in a sentence. The fully automatic machine translation system can produce English translations of input Japanese sentences, by making word to word translation, phrase to phrase structural translation, clause to clause structural translation, etc. at once within a second. User has no chance to see what happens in a translation system, so she gets lost at a glance of the meaningless sequences of word fragments and gets embarrassed in front of pairs of sound Japanese sentence and miserable English output in case of failure in particular. She, for example, a student in a college, even if she would like to have a sort of help for English composition, never try to use a MT as a tool again, after she incidentally has had an experience to sit beside those kinds of unconceivable screen outputs. A person like her usually has a fundamental grammatical and lexical knowledge for reading a textbook of English composition. She needs to be able to imagine how computer works and what it produces, before she starts to learn the operations.

TWP does have a very limited and robust capability of automatic translation with a variety of accessories including a reliable text book function, an useful dictionary function, an easily accessible wording examples database, etc. as well. It has been designed for a help for Japanese who has basic skill of writing English.

Some of those persons(Group A) who compose English need a very slight help like English dictionary lookup facility, some others(Group B) need a help like various consulting supports of bilingual dictionary, word sense definitions, translation examples, English grammar text books etc., the others(Group C) need a perfect help. TWP helps Group A and Group B mainly who work on Japanese word processor and English word processor for English production. They need very cheap, timely, comfortable support, and do not want perfect, expensive support necessarily.

A comfortable support in English composition on word processors includes the followings at least.

1. limited but sound translation facility if available,
2. a variety of word, bilingual, idiom, and wording dictionary look-up facilities,
3. a variety of translation example consultation supports,

There are many resources like the above available in the market and appears the following keen interests of users whenever they use them.

4. No irritation waiting for response: very quick response to every key or mouse strokes,
5. No eyesoreful data on a screen, because capacity of human eyes is very much limited,
6. Necessary information should be on a screen before user mentions it,
7. Easy recovery to the preceding screen from any unexpected result on a screen with no anxiety,
8. System controllability based on system transparency.

The point 1. of simple and fast translation capability affords to the point 4. and the point 8. transparency. The points 2. and 3. very simple and fast executable retrieving function support the point 4. the point 7. and very much the point 8. The points 4. through 8. are explicated here from now on.

Whenever TWP finds Japanese word end, phrase end or clause end, the translation equivalent is displayed on a screen line for inputting Japanese. J/E conversion needs scope determination of phrases and clauses. The TWP iterates scope determination and conversion very explicitly on a screen for realising traceability and transparency of translation.

Two issues are left necessary to explain further.

1. How to determine a scope of conversion ?
2. How to find an appropriate type of conversion ?

The figure 2 of intuitive image of J/E conversion depicts the two issues more clearly.

4.2 How to align a sequence of conversion step ?

The scope determination for conversion is one of key issues to specify comfortable user interface. The conversion should be carried out successfully at any case. Otherwise, users cannot rely on the conversion support and have anxiety of operation. The sequence of conversion steps should also be traceable within a limited human's eye capacity, that is to say, which conversion scope selected by TWP should be able to guess, and what sequence of conversions also should be foreseeable. User can catch up with every process very easily if user could expect every kind of change on a screen. TWP's idea to resolve user's anxiety comes from the above observation, to afford to the items 4. and 8. above greatly.

There are two reasonable scope interpretations for J/E conversion on the screen1 where a cursor stays just after a Japanese string 私は論文を which is composed of two Japanese phrases (bunsetsu)s,[私:I, は: Subject-marker,"I is the head of this noun phrase "bunsetsu],[論文: paper, を:Object-marker,"paper is the head of this noun phrase" bunsetsu]. One of the

```

Screen1 私は論文を
Word   watashi/wa /ronbun/wo
comment I      /SM /paper /OM
↓           ;Conversion the scope of "ronbun"

Screen2 私は essay を ;translation {essay, paper, ... }
Word   watashi/wa /ronbun/wo
comment I      /SM /paper /OM
↓           ;interactive pull-down menu selection
           of an alternative "paper"

Screen3 私は paper を書く。
Word   watashi/wa /ronbun/wo /kaku
comment I      /SM /paper /OM /write
↓           ;conversion of the scope of
           「私は paper を書く」

Screen4 I write a paper.
           ;English word order has been produced
           by simple sentence(clause) conversion

```

SM: subject-marker

OM: object-marker

Figure 2. Stepwise Japanese-English conversion

alternative scopes is "私は論文を C(cursor)", and the other is "論文を C".

TWP adopts such a simple strategy as a scope for conversion is the biggest logical unit of words, phrases, and clauses just adjacent to C(cursor) on the left part of input line. This strategy comes from Japanese intrinsic linguistic structure called head final.

Whenever an input is done, it is transferred to morphological processing to find logical units like word, phrase, and clause. A logical unit is handed to J/E conversion function to produce English equivalent. Japanese logical unit has a head word in the right part of it usually so that every end of Japanese unit is guessed much easier than that of English. Fortunately popular and robust algorithm of detecting end of unit is widely known and it looks ahead one word or one phrase to decide whether string on the left of cursor is on the end or not. In practice, TWP may recognize the end of unit more easily by hearing unit end signals from Kana-kanji conversion stroke as well.

TWP can guess scope of logical unit with high precision accordingly. The screen1 and screen2 in the figure 3 shows a selected conversion scope is the left smallest phrase(bunsetsu) which is usually signaled by kana-kanji conversion key stroke.

4.3 Simulated Mode of Automatic J/E Conversion

The "look ahead" algorithm detecting conversion scope by finding logical unit described above is also usable in another way. An already guessed logical unit

adjacent to a string on a cursor is finally recognized as a logical unit automatically if the string on the cursor could be guessed as a logical unit. This is called a "delayed" type of look ahead algorithm to finding logical bunsetsu units, which frees users from inputting logical unit end signals when input is displayed from left to right on a screen line. This algorithm permits an automatic detection of logical units for conversion. The figure three depicts clearly an implementation of automatic iterative J/E conversion of TWP.

The Japanese input facility is built in TWP automatically fixes the preceding bunsetsu (smallest phrase) just at the same time as a string adjacent to C(cursor) may be guessed a bunsetsu. The Screen1 bunsetsu candidate "彼は" with a cursor just after a bunsetsu end marker candidate は is not transferred to J/E conversion here. The "彼は" bunsetsu candidate is determined as a bunsetsu for conversion unit and is transferred to J/E conversion at the same time as a bunsetsu candidate "論文を" is guessed on the screen2 on its right. The second bunsetsu candidate "論文を" is fixed and transferred to J/E conversion when the next verbal bunsetsu candidate "書いた" is guessed on the screen3. The Japanese period "." signals an end of verbal bunsetsu "書いた" and triggers J/E conversion of it on screen 4. Here are three bunsetsus(smallest phrases). TWP recognizes easily a sentence, all of phrase head words of which have been converted into English, and transfers three phrases to J/E conversion altogether to produce the English expression on screen5. The product is well ordered and well translated with the help of simple translation function.

A Japanese and Alphabet mixed display format of intermediate conversion is a convention for user's tracing intermediate steps of J/E conversion easily on a screen. Japanese symbols in a mixed display convention express Japanese functional words like surface case-marker, conjunct particles, and auxiliary verbs. They are key information to determine grammatical meaning of head word in a sentence, and to execute structural J/E conversion, which should be transferred to translation component. That is to say, functional words expressed by Japanese symbols help Japanese users to keep grammatical structure of source Japanese sentence through English Composition on TWP. Those like case markers determine English simple sentence structure with case structure mapping scheme in a J/E bilingual dictionary corresponding to a head verb of a clause.

4.4 Screenplay examples of TWP

The figure 4(a) through the figure 4(c) demonstrate TWP's J/E translation support primitives. The first explicates traceability and transparency of TWP J/E conversion. The second exemplifies an easy and cheap cancel protocol to achieve user's contentment. The last shows an examples of dictionary consultation pro-

```

Screen1 彼は C
Word    kare/wa /C
Comment He /SM /C
↓
Screen2  he は論文を C
        ;論文 {ronbun} has plural translations
        {paper, paper, ...}
Word    He /wa /ronbun/wo /C
Comment He /SM /paper /OM /C
↓ automatic conversion

Screen3  he は essay を書いた C
Word    He /wa /paper /wo /kaita /C
Comment He /SM /paper /OM /write /C
↓
Screen4 he は essay を write た。 C
↓
Screen5 He wrote a paper.C

彼:kare:he
は:wa:subject-marker
論文:ronbun:paper/essay/article...
を:wo:object-marker
書いた:kaita:wrote
SM:subject-marker
OM:object-marker
C:cursor

```

Figure 3. TWP J/E automatic conversion step

cedure on TWP.

The figure 4(a) traces steps of real responses to Japanese input "このシステムによって / 論文を / 作成することが / できます /。", where "/" marks bunsetsu ends. This figure is composed of the five top lines of 5 screen copies from 1) down to 5). The forth line has three converted bunsetsus "this system," "paper", and "make" and one not converted bunsetsu "できます。" The bunsetsu "作成することが" is converted into the mixed format of "make することが" when and after inputting the forth bunsetsu is completed.

The figure 4(b) shows a cancel and undo protocol of TWP by listing five top lines of consecutive five screens from 1) to 5). TWP backspace key brings a status 1) to a status 2). After moving a cursor to "論文" in status 3), an user can edit to have "英文" instead of "論文" in a status 4). She can get an J/E converted expression of a sentence with "英文":English in a position of "論文":essay only moving a cursor to a sentence end position.

The figure 4(c) is composed of two screens. The screen 2) with three small windows open depicts what is displayed when dictionary consultation facilities are invoked. Under a mode of dictionary consultation, moving a cursor to any word either Japanese or English triggers dictionary look up with the word as a

key. The pull-down window w1 is opened to hold conversion dictionary contents of "essay" under a cursor. The window w2 is opened with "essay" as a key in the pull-down window w1 for further information of the word "essay". The window w3 has been opened with "paper as a key in w1. Lexical information on w2 and w3 comes from conventional bilingual dictionaries. Any word in w2 or w3 can be selected for a key of dictionary look-up.

5 Conclusion

TWP brings about both users' contentment with functions that machine translation has failed to realize, and large cost reduction for sentence generation in target language that simple dictionary consultation facility has not fulfilled. No other system has ever been presented to satisfy two of them at the same time.

TWP users reserve the means of cancelling and resolving the errors that the TWP or users created. TWP makes users feel content with processes on a screen by confirming every step of translation with their eyes. These help them to use it so easy with no anxiety.

References

- [1] Akamine, et.al., "English Composition Support Tool", Proc. of 43rd conference of IPSJ, pp.205-206, 1991
- [2] Aizawa, et.al., "Foreign Language Word Processor using Bi-directional Translation System", Proc. of 44th conference of IPSJ, pp.147-148, 1991.
- [3] Akamine, et.al., "Interface Protocol for interactive English Production on Japanese Word Processor", Proc. of 44th conference of IPSJ, pp.261-262, 1992.
- [4] Akamine, et.al., "Form Kana-Kanji Conversion to Kana-English Conversion", Proc. of 46th conference of IPSJ, pp.119-120, 1993.
- [5] Satoh, et.al., "User Controllable Translation based on Complex Sentence Translation Pattern", Proc. of 46th conference of IPSJ, pp.121-122, 1993.
- [6] Ando, et.al., "Translation with Lexical Learning Capability", Proc. of 46th conference of IPSJ, pp.123-124, 1993.
- [7] Nomura, et.al., "Japanese Word Processor with Word, Phrase and Sentence Handling Capability", NLP-SIG report of IPSJ, 1991.
- [8] Kamei, et.al., "Lexical Discourse Grammar", NLC-SIG report of Society of Electronics, Information and Communication of Japan, 1986.
- [9] Doi, et.al., "Lexical Discourse Grammar applied to Inter-clausal Parsing (II)", NLC-SIG report of Society of Electronics, Information and Communication of Japan, 1991.
- [10] Tsutsumi, et.al., "Editor for MT with Simple Learning Mechanism", Proc. of 33rd conference of IPSJ, pp.1771-1772, 1986.

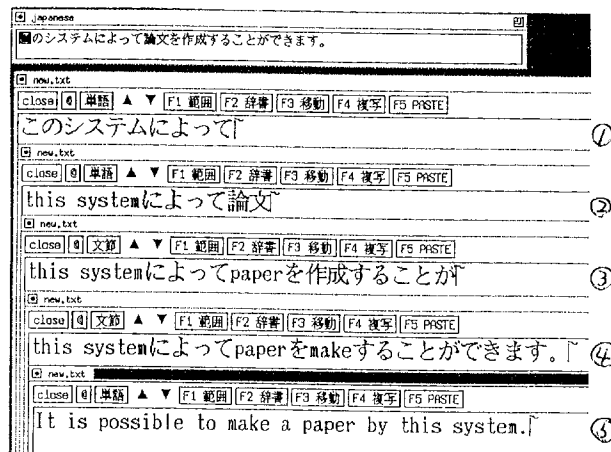


Figure 4(a). Traceability and Transparency of TWP

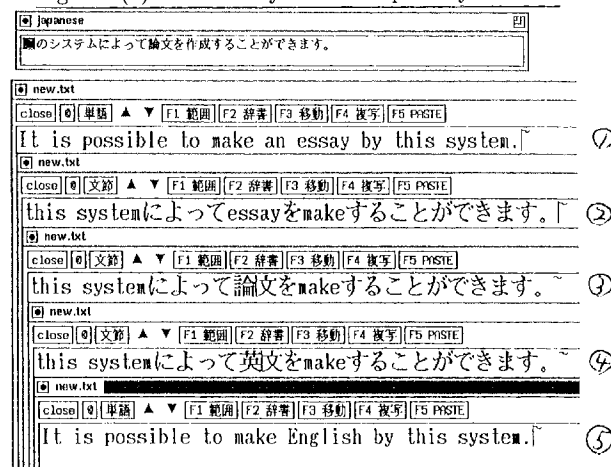


Figure 4(b). Easy and Cheap Cancel Protocol

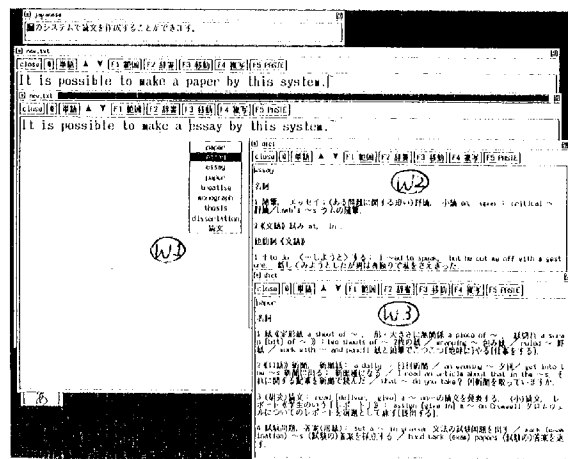


Figure 4(c). Dictionary Consultation Procedure