[From: IFTT'89, April 26-28, Oiso Prince Hotel, Japan]

What Should a Translation Work Bench Be like ?

Shin-ya AMANO, Kimihito TAKEDA, Koichi HASEBE

R & D Center, TOSHIBA Corporation

1. Introduction

The present machine translation objective is overall efficiency improvement, not full automation, of the translation process. Both makers and users were not clearly aware of the fact, or at least did not succeed in designing such systems, before 1980s. This was one of the reasons machine translation systems were not accepted in the office.

Even now we easily find people who think machine translation is fully automatic. Operational machine translation systems cannot be as simple as systems which read sentences and display their translation on a screen. These kinds of machine translation devices are on sale as a "machine translation system," but they are not worth being called a system, because they do not organize the translation process from a systematic point of view.

Operational translation systems are evaluated by their cost/performance covering the overall translation process, from inputting source language text to outputting target language text. There are several time factors, which could be reduced with a well-designed translation work bench, in the total translation process.

This paper reviews the time factors and how translation work benchs should be to reduce the time required in each process.

2. A typical Machine Translation Process

To clarify the above time factors a typical machine translation process is shown in Fig. 1.

[1] Inputting source language text

When text is given in magnetic medea, this process can be omitted.

[1-1] Reading text with an OCR

[1-2] Pre-editing

[1-3] Spelling check

Text is read through an OCR and

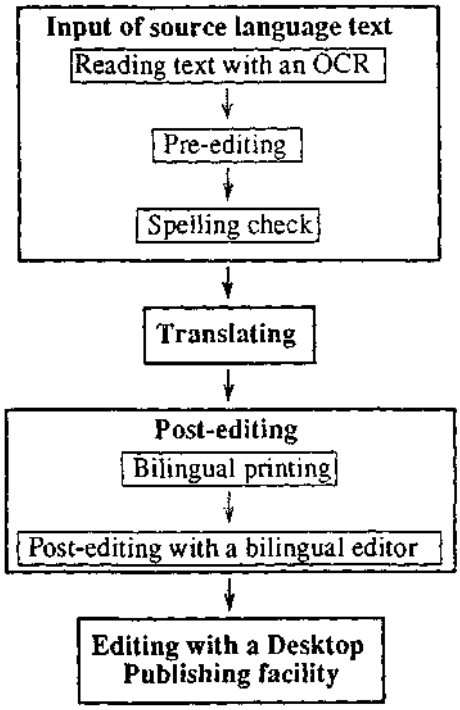


Figure 1. Typical MT process

some pre-editing should be carried out to delete unnecessary input data, such as headers, footers, page numbers, and parts of tables. Spelling check is necessary to correct OCR errors and to search for words which are not in the dictionary.

For machines, these two cases are the same. Machines consult their dictionary to find misread letters. If some word is not found in the dictionary, the word probably includes misread letters or it is an undefined word.
[2] Translating

Text is translated in batch mode.

[3] Post-editing

[3-1] Bilingual printing

Texts of source language and target language are printed in a sentence-to-sentence format. A human editor corrects the translation, referring to the source language.
[3-2] Post-editing with a bilingual editor

Source and target language texts are displayed with a bilingual editor in the same form as the above bilingual print. The correction is input with the bilingual editor.

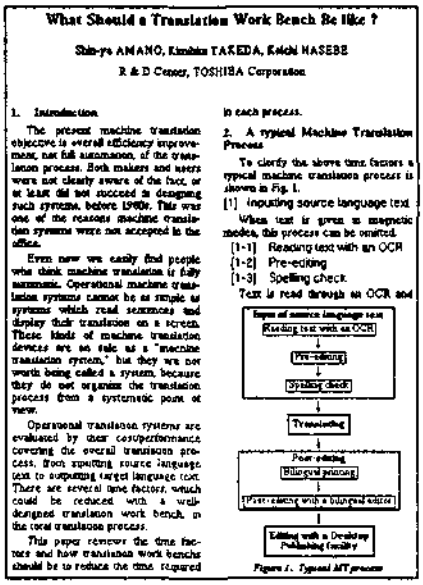
[4] Editing with Desktop Publishing facility

Figures, tables, photographs, etc. are input and some sophisticated editing can be accomplished with this facility.

3. Necessary translation work bench factions

3. 1. Text input

If source language text is not supplied in magnetic media, OCRs are necessary in machine translation systems to decrease the total translation process time. Functions and error rate are important OCRs factors


What Should a Translation Work Bench Be like ?

Shin-ya AMANO, Kimihito TAKEDA, Eiichi HASEBE

R & D Center, TOSHIBA Corporation

1. Introduction

The present machine translation objective is overall efficiency improvement, not full automation, of the translation process. Both makers and users were not clearly aware of the fact, or at least did not succeed in designing such systems, before 1980s. This was one of the reasons machine translation systems were not accepted in the office.

Even now we easily find people who think machine translation is fully automatic. Operational machine translation systems cannot be as simple as systems which read sentences and display their translation on a screen. These kinds of machine translation devices are no more than a "machine translation system," but they are not worth being called a system, because they do not organize the translation process from a systematic point of view.

Operational translation systems are evaluated by their cost/performance covering the overall translation process, from inputting source language text to outputting target language text. There are several time factors, which could be reduced with a well-designed translation work bench, in the total translation process.

This paper reviews the time factors and how translation work benches should be to reduce the time required in each process.

2. A typical Machine Translation Process

To clarify the above time factors a typical machine translation process is shown in Fig. 1.

[1] Inputting source language text

When text is given in magnetic media, this process can be omitted.

[1-1] Reading text with an OCR

[1-2] Pre-editing

[1-3] Spelling check

Text is read through an OCR and

Input of source language text
Reading text with an OCR
Pre-editing
Spelling check
Translating
Post-editing
Bilingual printing
Post-editing with a bilingual editor
Editing with a Desktop Publishing facility

Figure 1. Typical MT process

Figure 2. An exsample of a text

Any text usually has a complicated form, as shown in Fig. 2. It will contain tables, figures, headers, and footers, besides the body. Moreover it will employ multi-column format. These factors adversely affect OCR adoption, if it does not have functions to cope with them. In the worst case, manual input would be faster than OCR input.

Error rate is a critical OCR condition. In practical use, the word recognition error rate does not go down under 5%. A practical rate is 5-10%. This rather high rate does not just result in word reading error only. In text, there is a lot of coinage. It raises the virtual error rate, though OCRs read letters correctly. To cope with error, spelling checkers are essential.

An automatic paper feeder is also necessary for massive text input.

3. 2. Translating

Translation will be implemented in batch mode to achieve rapid translation. The interactive mode is not practical for massive translation. There are two keypoints in regard to cost / performance in this stage; translation speed and quality.

Needless to say, translation speed is a direct time factor. It should be as fast as possible. Translation quality is an indirect time factor. Its quantitative estimation as a time factor is realized in the post-editing stage.

3.3. Post-editing

Time necessary for post-editing consists of three factors; translation quality, post-editor man-machine interface quality, and human editor skill.

Translation quality depends on linguistic factors, such as a parser and a generator, and user-specific factors, such as lexicon and style. The latter is especially important for operational systems. If they are not customized, every sentence will be post-edited for trivialities in an expression, causing cost / performance to be very bad.

4. Conclusion

An example of the translation work bench is shown in Fig. 3, where rough estimation of machine translation and human translation is also shown.

Cost/performance for machine translation systems depends largely on OCR and post editing speed. Factors which affect the speed are ;

OCRs:
- Functions treating complicated text
- Error rate

Post-editing:
- Parsers and generators
- Customizing tools

This shows that not only linguistic factors but peripheral factors are important for achieving operational machine translation systems.

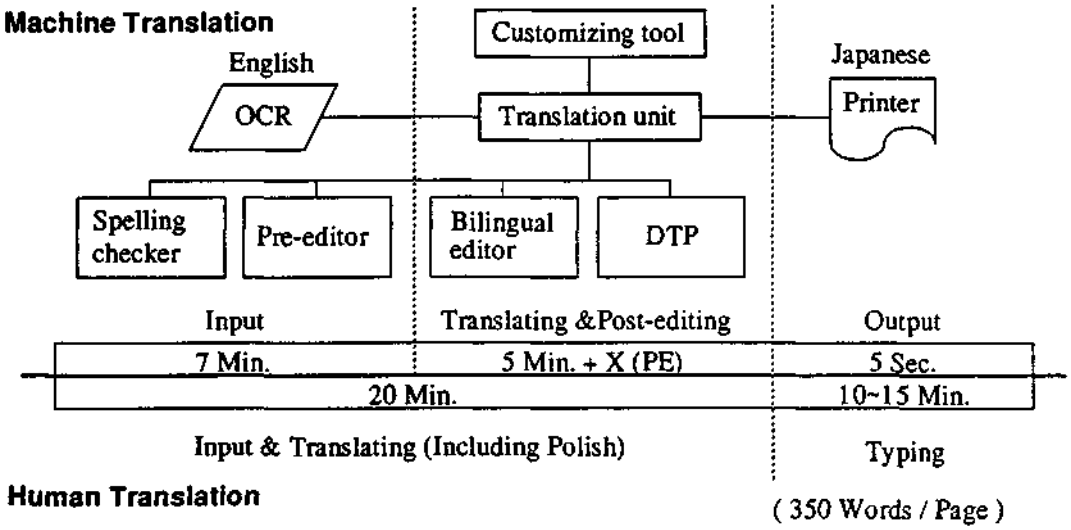


Figure 3. A TWB example and rough MT & HT estimation