

ALTFLASH: A Japanese-to-English Machine Translation System for Market Flash Reports

Hajime Uchino

NTT Cyber Solutions
Laboratories
uchino@soy.kecl.ntt.co.jp

Yoshifumi Ooyama

NTT Communication Science
Laboratories
ooyama@cslab.kecl.ntt.co.jp

Osamu Furuse

NTT Cyber Solutions
Laboratories
furuse@soy.kecl.ntt.co.jp

Abstract

We have developed a Japanese-to-English machine translation system for market flash reports called ALTFLASH. ALTFLASH is a hybrid translation system based on a combination of rule-based translation and template-based translation systems. The experimental results were that the system could achieve good translation for 90% of source sentences (70% of articles) in reports on the foreign section of the Tokyo Stock Exchange.

In addition, we focused on account settlement flashes, which formed fixed patterns, and developed a new system to translate them. This system has been installed, by Nihon Keizai Shimbun (Nikkei) in March 1998 in their English translation service for news flashes on settlements of accounts. It is a fully automatic translation system that enables news flashes to be broadcast to the world without requiring human intervention.

1 Introduction

Numerous systems and applications of machine translation have been developed. These include successful systems in operational use such as TAUM-METEO, a weather report translation system used by the Canadian Meteorological Centre in Montreal, and SYSTRAN, a large technical document translation system used by the US Air Force and General Motors of Canada. However, there are a lot of technical problems when translating from Japanese to English. Finally machine translation systems came to use for preliminary translation when translating large volumes of documents.

We have studied practical machine translation systems for industrial and economic information. In other languages, it exists some systems proposed to translate business news such as a translation system for Wire-Service economic news from English to Japanese (Aizawa 97) and a translation system for business news captions from English to German (Nyberg 97).

We have developed a Japanese-to-English multi-engine translation system for market flash reports called ALTFLASH, and confirmed its practicality for Japanese to English translation. And we focused on

account settlement flashes, which are basically formed in fixed patterns, and developed a new system to translate them.

This system has been installed by Nihon Keizai Shimbun (Nikkei) in March 1998, in the English translation service for news flashes on settlements of accounts. This is a fully automatic translation system that enables news flashes to be broadcast to the world without requiring human intervention.

This paper describes the characteristics of market flash reports in Section 2, gives an outline of the market flash translation system in Section 3, and provides experimental results along with practical applications of the settlement of accounts flash translation system in Section 4.

2 Characteristics of Market Flash Reports

In Market flash reports, there are many expressions that are difficult to translate with a conventional machine translation system. Such expressions include technical terms or domain-specific expressions. Furthermore, similar expressions are used often. With these factors in mind, we constructed the translation system.

For our experiment, we used articles on the foreign section of the Tokyo Stock Exchange, the Adjusted Stock Price Average in Osaka and Tokyo CBs during 14 weeks (June-Sept. 1995). These articles were taken from the Telecom Database of Nikkei, and paired Japanese with English by using alignment system (Takahashi 97). A sample of article is shown in Fig.1.

Each article consists of a headline, a "market summary" section (one sentence), a "general conditions and backgrounds" section (one to three sentences), and a "movement of individual issues" section (two or three sentences).

English headlines always have descriptions of market movement, but Japanese headlines often lack them.

The "market summary" section often include similar expressions, and so template-based translation seems appropriate. Moreover the "Market summary" section in English uses the day of week that corresponds to the date, so processing to insert it is necessary.

As for the "general conditions and backgrounds" section, sentences are long and causal relations are

Japanese Headline	◇東証外国部・大引け
Body of Article	【NQN】欧州株の軟調を映しさえない展開。売買高は概算10万株。売買が成立したのは36銘柄(値付き率53.7%)で、このうち値上がり 5、値下がり 18、変わらず 2、比較できず11だった。イーライリリが安く、VW、IBM、BTも下げ、ドイツ銀行もさえない。半面、モトローラが買われ、TI、アップルも高い。
English Headline	*Tokyo Foreign Stocks CIs: Mostly down on weaker European shares
Body of Article	(NQN) Foreign stocks declined Friday. An estimated 100,000 shares changed hands. Among the 36 issues traded, only five rose, 18 dropped, and two were unchanged. Comparisons for the other 11 were unavailable because they did not trade on the previous day. Eli Lilly, Volkswagen and IBM softened. In contrast, Motorola, Texas Instruments and Apple Computer gained.

Figure 1: Sample Article (November 8, 1996)

often described. We aim to improve accuracy by dividing long sentences.

Sentences of "Movements of individual issues" section have extremely fixed forms, so template-based translation is suitable. Processing for parenthetical expressions is indispensable.

3 Structure of Translation System for Market Flash Reports

After giving consideration to the above features of market flash articles, constructed the translation system shown in Fig.2. Each translate engine is running in parallel, and the results of the translation are assembled after completing every sentence of an article.

Pre-processing system

- The character code of the received article is converted, and the data is edited for sending to the translation engines.
- The headline and first sentence of the body are sent to the headline translation engine, and all sentences of body are sent to other translation engines.

- The parenthetical expressions are classified, and if an expression is a translation target, it is isolated as a separate sentence and its information is recorded for later filling in.

Headline Translation

A Japanese headline is composed of section (ex. "Tokyo Foreign Stocks"), session (ex. "Opg", "Mngcls"), background (ex. "N straight days" is background in "Down for N straight days"), movement (ex. "drop", "continue rising"), and comment parts (right of bar, often lacking).

These components are directly transferred to English. If a Japanese headline does not have a movement part as in the headline of Fig.1, a keyword is obtained from the first sentence of the body.

Template-Based Translation

A Direct transfer method is effective in translating the sentences that have fixed patterns frequently used.

When translating the Market flash articles by using templates, the variable sections are mainly "company name" and "number". To use simple literal extractions of variable sections would erroneously extract company names and numbers because company names are frequently shortened in market reports. In addition,

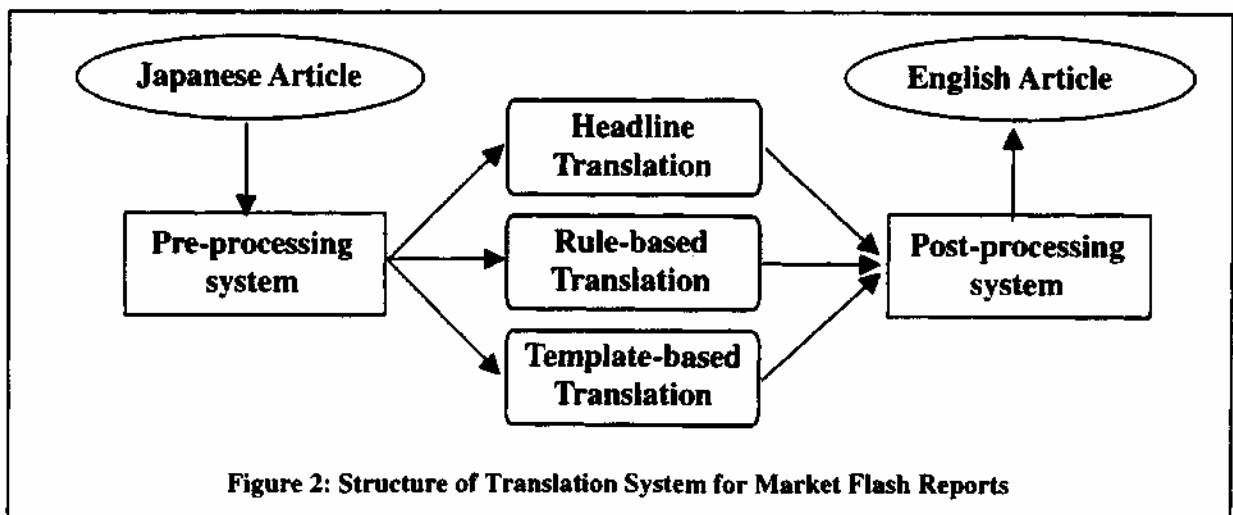


Figure 2: Structure of Translation System for Market Flash Reports

considering noun phrases as variable sections would expand the applicable area for templates. Thus we use morphological analysis to segment words.

After the morphological analysis, variable sections are decided by matching segmented sentences to rules. Simple words are translated by dictionaries, and noun phrases and noun compounds are translated by a rule-based system.

Rule-based Translation

This module is realized through the use of the ALT-J/E system (Ikehara 1991). The ALT-J/E system has a very large dictionary. The dictionary classifies the meanings of words into 3000 categories, based on which it stores 400,000 word definitions and 16,000 patterns of semantic structures.

For the particularity of articles, we did not construct the translation stage using specific grammar (Aizawa 96) but added domain-specific rules and dictionaries. The additional dictionaries are a name dictionary (2,000 words) and a financial term dictionary (6,000 words).

We aim to improve accuracy by dividing long sentences in the general conditions and backgrounds.

Post-processing system

This process assembles the translation result of an article from the results of each translation engine. The translation result of each sentence is scored on the following basis, and the result with a high score was selected.

- 1) A basic score of 10 is attributed to the results coming from the template-based system. For each internal part of the translation that results from the use of the rule-based system, we subtract 1 point to this score.
- 2) A basic score of 6 is given to the results coming from the rule-based system; 3 points are subtracted for each non-translated word.

Because the templates easily match the sentence in the second half of an article, the translation result was assembled before the completion of rule-based translation in many articles.

After assembling articles, the parenthetical expressions that were isolated by pre-processing are

filled in. Finally, the addition day of the week and the deletion of lengthy words are done.

4 System Performance

4.1 Experimental Results on Market Flash Reports

The translation of the Japanese sentence in Fig.1 by this system is shown below. The underlined sentences were translated by a template-based system.

Tokyo Foreign Stocks Cls: Lower

NQN) Foreign stocks were clear reflected by the bearishness of European stocks Friday. Trading volume was estimated at 100,000 shares. Among the 36 issues changing hands, five rose. 18 fell, two remained unchanged, and 11 were incomparable. (53.7% of all listed issues changed hands). Eli Lilly is lower and Volkswagen, IBM and British Telecom also fell and Deutsche Bank also faltered. In contrast, Motorola was bought along with Texas Instruments and Apple Computer.

The experimental results are described as follows.

Headline translation system:

77 headlines of 80 articles (Sept. 1995) translated accurately.

Template-Based translation system

The progress of making template rules for Tokyo foreign stocks are shown in Table 1.

In this field, 80% of sentences were translated by template (60% for Osaka Stocks, 45% for Tokyo CBs). Hit rates after making templates were nearly stable in any fields, and the rate of making templates also decreased gradually.

However, many sentences that describe "summer vacation" were not matched in August's articles, and also sentences that describe "midterm settlement" were not matched in September's. It is necessary to make templates for an entire year because there are other annual events.

Total evaluation

The translation results of sentences were evaluated into four grades, and the top two grades indicate success. If every sentence in an article is successful,

Table 1: Progress of making template rules for Tokyo Foreign Section

	Sentences	Hit rate 1	Hit rate 2	Templates made	total templates	Making rate
95.6	58	0.0%	89.7%	20	20	34.5%
95.7	221	35.3%	82.4%	42	62	19.0%
95.8	274	55.1%	85.8%	38	100	13.9%
95.9	236	61.1%	81.4%	29	129	12.3%

Hit rate 1 = number of hit sentences before making templates

/number of source sentences (Blind Test)

Hit rate 2 = number of hit sentences after making templates

/number of source sentences (Window Test)

Making rate = number of templates made/number of source sentences

the article is successful.

In these experiments on 80 articles, the success rate for sentences is 74% and the success of articles was only 23%. However in the Tokyo Foreign Stocks where templates were applied at high frequency, the success rate reached 90% for sentences and 70% for articles.

4.2 Practical use of the system

From the above experiments, there is a fair prospect of practical use of the system in limited fields. However post-editing is necessary with the present system because polished English is required for business services. Therefore, we focused on account settlement flashes, which are formed with very fixed patterns, and developed a new system in cooperation with Nikkei.

4.2.1 Modification of System

Structure of translation system for settlement flashes is shown in Fig.3. We modified the system as follows. To improve the translation speed, we removed the headline translation system and rule-based system from

the prototype system and added a direct transfer system. The direct transfer system does not analyze morphemes but uses simply extraction of variable sections. When failing to extract variable sections, a conventional template translation system that runs in parallel processes it. In post-processing, the result that returned early was chosen.

By applying the N-gram Method (Ikehara 96) to account settlement flashes for one year, all fixed patterns were extracted. We made about 200 template rules from these patterns and confirmed that the rules covered all patterns. Examples of template rules and sample translation are shown in Fig.4.

The pre-processing system has an article database to record correspondence relations between Japanese articles and English articles. If Japanese articles are modified, the system also modifies the corresponding English articles.

The monitoring system displays the translation results and warnings. If a company name that is not listed is used or there is a typing mistake in a Japanese flash, the translation system does not translate but sends a warning to the monitoring system (Fig.5).

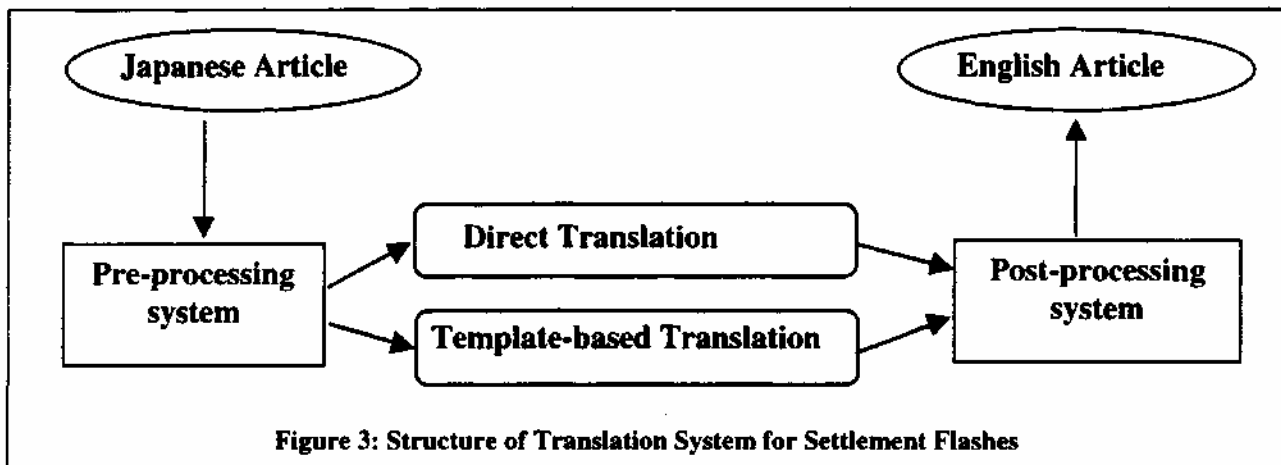


Figure 3: Structure of Translation System for Settlement Flashes

Samples of template rules

COMPANY /of/previous period/pre-tax profit / No.1 / yen
 [企業名]/の/ 前期 / 経常益 /[数1]/円



[COMPANY] FY PARENT PRETAX PFT Y [No1]

COMPANY /of/4-6 month period/operating red ink/ No.1 / yen/--/previous year/ same period /is / No.2 / yen
 [企業名]/の/ 4-6月期 / 経常赤字 /[数1]/円/--/ 前年 / 同期 /は/[数2]/円



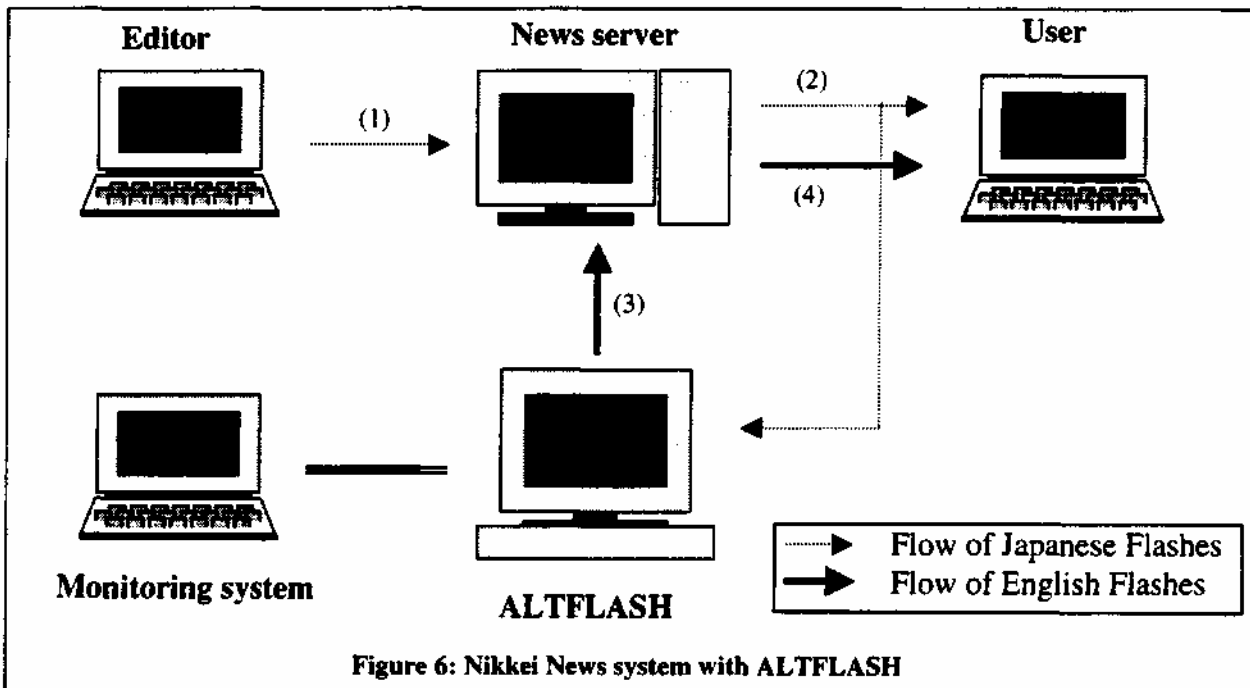
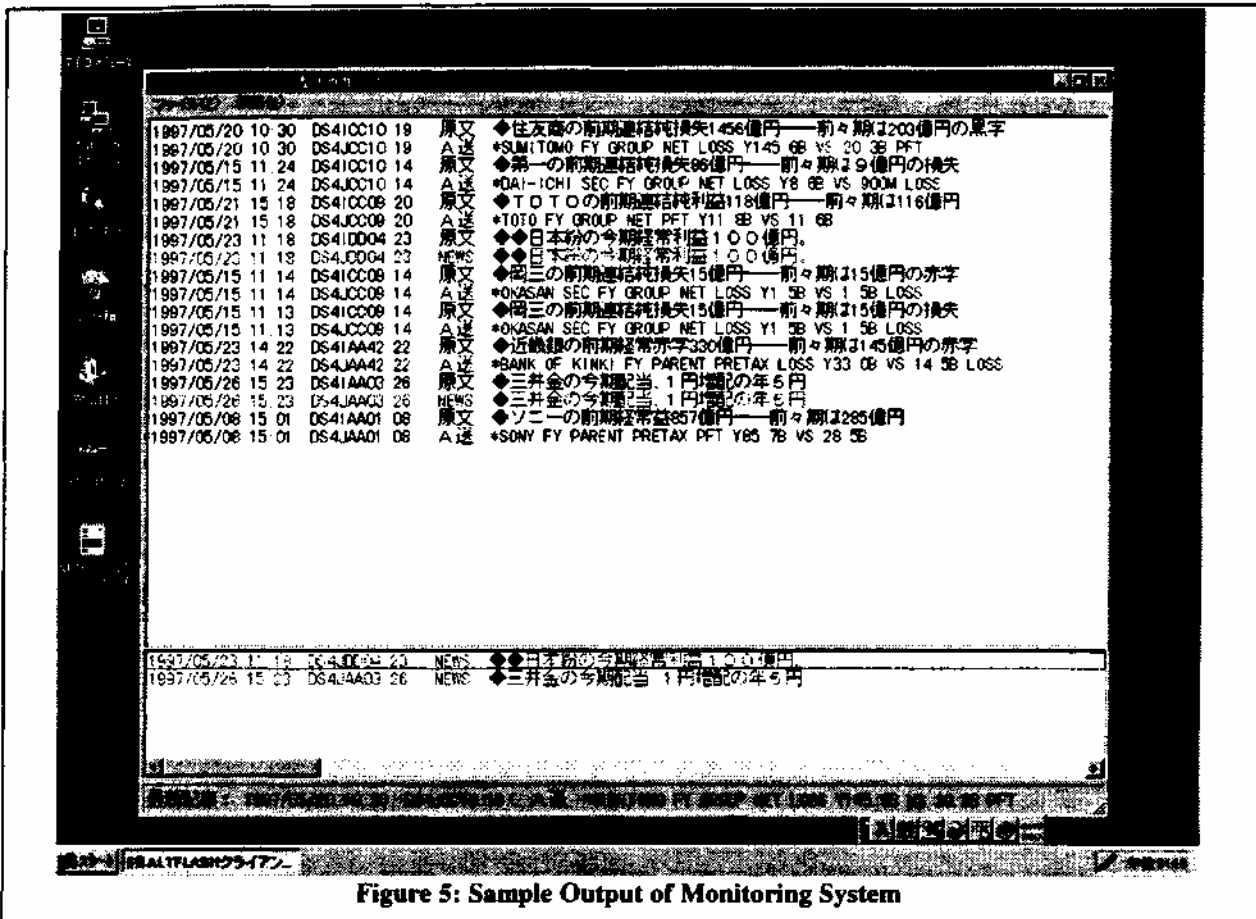
[COMPANY] SEES CURRENT FY PARENT PRETAX PFT Y [No1]

Sample of translation

松電工の前期経常益 5 1 2 億円

MATSUSHITA ELEC WORKS FY PARENT PRETAX PFT Y51.2B

Figure 4: Samples of Template Rules and Translation



4.2.2 Benefits of using this translation system

This system has been installed by Nikkei in March 1998 in their English translation service for news flashes on settlements of accounts (Fig.6). This is a fully automatic translation system that enables news

flashes to be broadcast to the world without requiring human intervention. The average time after delivering the Japanese flash until the English flash is broadcast is under one second. The benefits of using this system are as follows.

Reduction of work

Before ALTFLASH was introduced, the settlement of account reports announced by Japanese companies were first sent out in Japanese and then manually translated into English. Although settlement flashes are simple sentences, number errors or typing errors were not allowed, and it was tense work. Now freed from the burden of translation, human translators can concentrate on more sophisticated work.

Decrease in incorrect flash reports

When there was a mistake in a Japanese article, because a warning immediately alerts users to mistakes, incorrect reports have decreased. Of course, the English articles are considered perfect.

Shortening of time spent on dispatch

Firms' announcements of settlement of accounts concentrate in May and November. Sometimes over 200 companies announce in one day. When a large number of articles came out at once, the dispatch of English articles was delayed.

The introduction of this system, helped reducing the necessary time to achieve the broadcast of the translations in such a critical period.

5 Conclusion

In this paper, we described the design and the implementation of a machine translation system for market reports, from Japanese to English. The translation system for account settlement flashes is in operation now, and its expansion to articles of dividend information and earnings revisions is being considered.

We are also studying a prototype system to increase translation accuracy. Furthermore, we are planning to expand the system to other domains and to other languages.

References

- Nyberg E. and Mitamura T. (1997). "A Real-Time MT System for Translating Broadcast Captions". In Proc. of the 6th MT SUMMIT, pp. 51 -57
- Wakita E. (1998). "Development of Automatic Machine Translation System for Account Settlements Flash". In Newspaper Technology, No.165, pp.35-41 (in Japanese)
- Ikehara S., Shirai S. and Uchino H. (1996). "A Statistical Method for Extracting Uninterrupted and Interrupted Collocations from Very Large Corpora". In Proc. of COLING'96, pp.574-579
- Aizawa T., Katoh N. and Kamata M. (1996). "English-to-Japanese Machine Translation for Wire-Service Economic News", In Journal of Information Processing, Vol.37, No.6, pp.1041-1048 (in Japanese)
- Ikehara S., Shirai S., Yokoo A. and Nakaiwa H. (1991). "Toward an MT system without pre-editing - Effects of new method in ALT-J/E -". In Proc. of 3rd MT SUMMIT, pp. 101-106
- Uchino H., Shirai S., Ikehara S. and Shintami M (1996). "Automatic Extraction of Template Patterns Using n-gram with Tokens". In Technical report of IEICE, NLC 96-18, pp. 63-68 (in Japanese)
- Takahashi Y., Shirai S. and Bond F. (1997). "A Method of Automatically Aligning Japanese & English Newspaper Articles". In Proc. of NLPRS'97, pp.657-660