

Translation of News Headlines

Kenji Ono

Research and Development Center
Toshiba Corporation
Komukai-Toshiba-cho 1, Saiwai-ku, Kawasaki, 210-8582, Japan
email: kenji2.ono@toshiba.co.jp

Abstract

Machine-Translation of news headlines is difficult since the sentences are fragmentary and abbreviations and acronyms of proper names are frequently used. Another difficulty is that, since the headline comes at the top of a news article, the context information useful to disambiguate the sense of words and to determine their translation(target word) is not available. This paper proposes a new approach to translating English news headline. In this approach, the abbreviations and acronyms in the headlines are complemented with their coreference in the lead of the article. Moreover, the target word selection is performed by referring to the translation of similar news articles retrieved from a parallel corpus. In the experiment, 100 English headlines are translated into Japanese using a corpus containing 30,000 English-Japanese article pairs, resulting in a 17 % improvement in the target words and a 21 % improvement in the style of translation.

1 Introduction

Recently, commercial machine translation(MT) software for the PC has become a popular tool for Internet surfing. People often use it to read foreign online news pages in order to obtain the latest world information. In Japan, for example, the home pages of “The Washington Post”, “The Wall Street Journal”, and “Yahoo! News” are among the most widely read pages using Internet browsers embedded with English-to-Japanese(*E-to-J* hereafter) MT functions.

However, news headlines pose a great challenge to conventional MT systems since they are full of abbreviations and omissions and are written in a very short style.

To make the matter worse, the words that are often abbreviated are current, topical names which are rarely registered in the commercial MT system lexicons, since they are updated on a yearly basis at best. Even if abbreviated words are translated correctly, they are sometimes hard to understand for non-Western people who lack some cultural background knowledge and are possibly ignorant of well-known English proper nouns. For example, some Japanese may be at a loss to see the translation of word “Enron” in a headline. If the word

were “Enron Corp.”, then the translation would be different and hence be very informative for those who see the word “Enron” for the first time, since they understand “Enron” is not the name of a place or a person but the name of a company.

Another difficulty in translating news headlines is the lack of context information needed for target word (*tw* hereafter) selection. A conventional MT system utilizes information from previously translated sentences to select the best translation of polysemous nouns in the succeeding part of the document. This process is considered a part of Word Sense Disambiguation(WSD). For example, to translate the word “goal”, MT systems select “ゴール (soccer goal)” instead of “目的 (aim)” if the word “soccer” appears in the preceding sentences. However, since the headline is the first sentence of the document, context information for target word selection is not available at the time of translation.¹

Although recent improvement of PC hardware allows MT systems to look ahead at suc-

¹Typical commercial MT systems usually do not refer to succeeding sentences in order to achieve fast translation and to meet system requirements.

ceeding sentences before translation, WSD remains a big problem, since translation rules, like specifying the relation between “soccer”-“goal”, are ad-hoc and cannot be prepared systematically.

For these reasons, the translation of news headlines is quite difficult. However, their correct translation is crucial, since readers usually read only the headline to decide whether they will go on to read the body or not.

The rest of the paper is organized as follows. In Section 2, I will explain my approach in detail, and a headline translation system based on this idea is shown in Section 3. Section 4 illustrates the translation process of the system. I will present the result of the experiment in Section 5 and draw a conclusion in Section 6.

2 The Approach

2.1 Named Entity Extraction and Coreference Resolution

In news translation, it is important that proper nouns, including people’s names or company names, are translated correctly. But, in the headline, these Named Entities are often abbreviated or shortened. For example, the expression “President George W. Bush” is often replaced with “Bush”, and “Walt Disney Co.” is simply shortened to “Disney”. Without these prepositives and postpositives, it is very difficult for MT systems to disambiguate whether, say, “Bush” is a person’s name, a company’s name, or a plant with branches. Also acronyms like “Hong Kong Military Authorities” → “HKMA” have similar problems.

Fortunately, in news articles the full spelling can be usually found in the lead, especially in the first sentence of the lead. Therefore, it is expected that the translation of Named Entities in the headline will be improved if they are replaced with the corresponding expressions in the lead in advance of the translation.

To complement shortened Named Entities in headlines with full expressions in the lead, two techniques are needed: *Named Entity Extraction(NEE)* from the headline and lead, and *Coreference Resolution(CR)* between them. In (MUC6, 1995) and (MUC7, 1998) substantial studies on NEE and CR have done. The best performance on NEE task in MUC7 was achieved by (Krupka and Hausman, 98)and

(Andrei Mikheev and Moens, 98) whose approach were dictionary and rule-pattern based. As for CR, recently (Ng and Cardie, 02a) and (Ng and Cardie, 02b) devised a supervised learning method for CR which outperforms the best CR results of MUC6 and MUC7.

I translated 300 English headlines into Japanese and picked out the Named Entities which were incorrectly or poorly translated and whose translation would be improved by the CR with the lead. The analysis lead me to conclude that the pattern of Named Entity expression is very specific and, like (Andrei Mikheev and Moens, 98), the rule-based approach which extracts Named Entities based on the pattern of word capitalization, collocation and Part of Speech(POS) would be promising for NEE.

On the other hand, the analysis showed that the CR task between the headline and the lead does not fall under the general CR framework, since the use of the definite article *the* in the headline, anaphoric expressions like pronouns, and rephrasing in the lead, which are the primary clues for general CR task, are hardly observed. It is more like a word-matching task between Named Entities in the headline and those in the lead.

In accordance with these observations, I adopt the rule-based approach for NEE and a simple linking method for CR. The details will be given in Section 3 where NEE is performed as a part of the CR process.

2.2 Target Word Extraction

In order to select appropriate target words for polysemous words, several studies on the use of context information have been done.

(Ooi and Tsunoda, 96) proposes to disambiguate nouns in news headlines by context information from the body, utilizing WordNet and LDOCE. However, for the task of target word selection for news headlines, it is supposed that general word sense knowledge from databases such as WordNet is too general, and the disambiguation result obtained through the mapping of WordNet is likely to be a basic, simple word that rarely appears in news articles.

In (Yoshimura et al., 97) the use of subject area information is proposed for web page translation. To validate the usefulness of subject area information for news article translation, I performed a preliminary observation in which the

target word of a polysemous word in the English article is chosen from the Japanese articles of the same subject area. The articles are retrieved from *Nikkei Shimbun* corpus, a Japanese large-scale monolingual corpus which contains 200,000 news articles/year classified with an 8-level thesaurus(subject area). In the experiment, the level2 thesaurus(151) and level3 thesaurus(1,474) are used. The result proved that for the translation of, say, “grade”, a subject area classification such as “education” is too general to determine which is preferable between “学年 (*gakunen*, a school year)” and “成績 (*seiseki*, school record, result of examination)”.

On the other hand, there are several studies on corpus-based Word Sense Disambiguation. (Kaji and Morimoto, 2002) and (Jie and Yuhang, 2000) applied a noisy or comparable bilingual corpora for the WSD of E-to-J and E-to-C(Chinese) translation. These studies are statistics-based, which implies that a disambiguation rule should be extracted from a certain quantity of text in order to be statistically significant. This quantity of text can be considered to be comparable with the articles in the *subject area* described above, and it follows that a corpus whose size is as large as 200,000 documents is insufficient to statistically disambiguate and decide the target word of, say, “grade”.

Meanwhile, it was also observed in the experiment that good target words were obtained from articles of the same *event* classification, which is a finer-grained group than subject area classification. For example, to decide the target word of “share” between “シェア (*shea*, percentage of the market)” and “株 (*kabu*, company stock)”, subject areas such as “business” or “company M&A” are too coarse, while classification levels such as “company share-listing” or “market analysis” provides good disambiguation, although the number of documents of an event like that are quite small and seem to be insufficient for the stochastic approaches.

In accordance with these observations, I decided not to make use of statistical approaches and decided instead to utilize a parallel corpus from which the English articles which are about the same event are retrieved, and then the adequate target words are extracted from their cor-

responding Japanese articles. The details will be given in Section 3.

3 The Headline Translation System

Figure 1 outlines my headline translation system according to the approach described in Section 2. As shown in the figure, the system is composed of the **preprocessing unit**, the **coreference resolution unit(CR unit)**, the **similar article retrieval unit(SAR unit)** and the **target word extraction unit(TWE unit)**. The **translation unit** performs conventional machine translation, referring to the result of preceding units. The following subsections look into each unit in detail. For simplicity, in the rest of the paper I will refer these units as the **CR unit**, the **SAR unit** and the **TWE unit**, respectively.

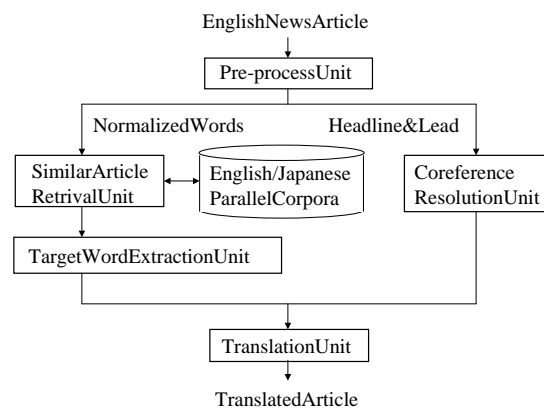


Figure 1: Overview of the Headline Translation System

3.1 Preprocessing Unit

First, the preprocessing unit analyses the input article(web news page) to detect the headline, the lead, and the rest of the body.² It also performs a morphological analysis, and the words in the article are normalized so as to match the entries of the system lexicon.³ Then, the headline and the lead are sent to the CR unit, and

²If there are more than one news articles on a Web page, the system divides them and repeats the whole procedure for each of the articles.

³Plural forms are converted to singular forms and inflection forms are uniformized to infinitive forms. Simple writing variations such as upper/lower cases are also uniformized.

the normalized words from the whole article are sent to the SAR unit.

3.2 Similar Article Retrieval Unit

The SAR unit retrieves similar English articles from a parallel corpus by a standard vector space model with cosine normalization.

The method of finding similar documents in the corpus can be considered as a kind of article alignment. (Collier et al., 98) and (Takahashi et al., 96) achieve high accuracy in news article alignment by confining the articles to be retrieved to those which have either the same date expressions, the same numeric expressions (amount of money, etc.), or the same proper names in the text.

However, for the purpose of extracting target words suitable for the translation of similar news articles, it is sufficient that retrieved articles treat *the same kind of events* (company share-listing, fire, airplane-crash, etc.), and they do not have to be related to *the same incident* as the query document. Requirements such as sharing date expressions or proper names inevitably limit the amount of retrieved documents, and it is likely that a sufficient number of target word candidates applicable for translation will not be extracted.

From this point of view, the SAR unit deliberately lowers the weights of date expressions, numerical expressions, and proper names in cosine normalization calculation.

3.3 Target Word Extraction Unit

the TWE unit scans the Japanese articles which correspond to the English articles retrieved by the SAR unit, and selects the best target word of a polysemous word in the English article by the following steps:

1. for any word(source word) in the English text, count the appearance of the word(target word) in Japanese text which the E-to-J MT system lexicon provides as the translation of that source word.
2. count the number of source word-target word pairs.⁴

⁴Actually, the number of pairs of Japanese source word - English target word are also counted by referring to the J-to-E MT system lexicon, and both counts are summed before step 3.

3. output the pair whose count is max.

This approach is straightforward in that it is not based on any sophisticated statistics and does not rely on any sentence-alignment or word-alignment processes which are known to work well for parallel corpora. The reason is that since the number of English-Japanese article pairs of the same event are small, as mentioned in Section 2, the number of source word-target word pairs are too small to be statistically treated.

Also, the preliminary experiment showed that the deterioration of the output, by not relying on sentence or word alignment, is rather small and will be compensated by the advantage of this approach, that it can be applied to noisy, comparable bilingual article pairs to which sentence or word alignment is not applicable.

3.4 Coreference Resolution Unit

The CR unit performs the extraction and coreference resolution of the Named Entities in the headline and lead by the following steps:

1. Extracts the word chunks (Named Entity candidates) from the headline and the first sentence of the lead, using rules over the series of words, POS, and a set of stop words. The rules also refer to the literal information, such as dashes, parentheses and capital letters.
2. Calculates the similarity score between a chunk in the headline and one in the lead. In calculating the similarity, the number of words that appear in both chunks plays a primal role. Acronyms such as “Hong Kong Military Authorities” → “HKMA” are also treated in this step.
3. For every chunk in the headline, outputs the chunk in the lead that has the most similarity to it.

The CR process mentioned above seems to be in favour of recall and against precision, since any phrase in the lead will be extracted if it contains the same proper word as in the headline. However, the loss of precision is expected to be small by the restriction that the Named Entities be extracted from only the headline and first sentence of the lead.

3.5 Translation Unit

The translation unit translates the headline referring to the result of the TWE unit and the CR unit. The phrases for which the CR unit has found full chunks in the lead are replaced prior to translation.

As for the words for which the TWE unit has detected the target words, the translation unit prefers them so long as grammatical or semantic conditions coded in the translation rules of the MT system are met. The extracted target words are also used in the translation of the rest of lead and the article body.

(Yoshimi and Sata, 99) and (Shirai et al., 97) analyze the style of headlines from a grammatical point of view and propose headline translation rules. Our system adopts headline-specific translation rules which treat *taigen-dome*, *kuten*, to-infinitives, present participles, and the words “see” and “say”. See Section 4 for the effect of these rules.

4 Illustration of the Translation Process

This section illustrates the transaction process of the implemented headline translation system.

The upper part of Figure 2 shows the headline and lead of an example news article. First, the CR unit performs coreference resolution between its headline and lead and obtains the following output:

“Shanghai New Asia” →
“Hotel operator Shanghai New Asia”

In this case, only one Named Entity is extracted.

Then, for this article, the SAR unit calculates the similarity of English articles in the parallel-corpus.⁵ The top 10 ranking of articles and their corresponding Japanese articles are shown in Figure 3. It is clear that the retrieved articles are all related to share-listing events, irrespective of the company’s name and the amount of money or shares involved. The TWE unit then extracts the source word - target word pairs from all these E-to-J article pairs that the SAR unit retrieved.

Figure 4 shows the examples of source word - target word pairs that the TWE unit has ex-

⁵All words in the article, including its body, are used in the calculation.

tracted from 10 article pairs in Figure 3. In this figure, “上場する” is extracted as the translation of “list”, compared to the original target word “リストする”, where “リスト” is the *katakana* translation (literal or phonetic translation) of “list”.

The word “list” has many meanings, and different Japanese translations fit for each of them. For a basic polysemous word like this, unless some extra disambiguation information is available, conventional E-to-J MT systems apply the most “safe” translation. This word is called its *default target word*. A Japanese *katakana* word vaguely implies all senses of the original English word and is usually suitable for its default target word, but it is not always the best translation in the given context. For a news about share-listing event like this, the extracted target word “上場する” is the best translation of “list”.

On the other hand, the word “issue” has no *katakana*-translation and the default translation is “問題”, which is wrong in this case. In the share-listing domain, the best translation is “発行”, which is correctly extracted. As for the translation of “new” and “raise”, more official or formal translations are extracted which fit to the translation of news and hence contribute to the stylistic quality of translation.

The lower part of Figure 2 presents the translation result of the headline in comparison with that of a conventional MT system. As you see, “Shanghai New Asia” is translated to “ホテル経営者上海ニュー・アジア” due to the result of CR unit, which explicitly tells “Shanghai New Asia” is a hotel operator.⁶ The translation of “New Asia” changed from “新しいアジア”(semantic translation) to “ニュー・アジア”(katakana translation joined with centered dot), which is better for the translation of company name.

The effect of the extracted target word produced by TWE unit is found in the translation of “list”. Among the extracted translations in Figure 4, “上場する” is applied for the translation of “list”.

From the stylistic point of view, the translated sentence ends without “する。”,⁷ and the

⁶The best translation of “hotel operator” is “ホテル運営会社”(hotel-management-company) since “ホテル経営者” implies a person.

⁷so-called *taigen-dome*(the nominalization of verbal noun).

subject is marked up with “、 (*kuten* punctuation mark)” in stead of “は (*ha*, postpositional particle)” , This style is typical in Japanese news headlines and is brought by the translation rule described in Section 3.5.

The body and its translation of this article is omitted in Figure 2 but it has been observed that the application of extracted target words shown in Figure 4 improved the translation.

5 Evaluation

Our experimental headline translation system has been implemented⁸ based on our commercial E-to-J MT system.

The parallel corpus is composed of a clean-parallel corpus and a comparable corpus: approximately 20,000 Reuters articles of all genres in 1995-1996⁹ and their manual Japanese translations, and 11,000 world news articles and their translations which were collected from U.S. and Japanese websites and were article-aligned by the method described in (Collier et al., 98).

I applied the system to 100 Reuters articles and compared the translation results with those from the original MT system.

The average number of retrieved articles from the corpus is 3.2, but the standard deviation is 3.6, which means the number of articles which treat the same event varies depending on the event type. In fact, the number of articles which have more than 9 similar articles in the corpus is 16, while the number of articles for which no similar articles can be retrieved is 57. The average number of extracted target words is 17.0. The average number of extracted CRs is 1.0.

The following summarizes the improvement of translation as a whole.

target word improvement: In 26 translations the target words are better than the original, while 9 translations worsened. Among them, CR affected 24(better:17, worse:7) and TWE affected 11 (better:9, worse:2). Thus, 17% of the translation improved. The worsened cases of CR are attributed mainly to the wrong Named En-

tities or the wrong translation of Named Entities being extracted by CR.

style improvement: In 28 translations, the style is better than the original, while 7 translations worsened. Thus, 21% of the translations improved. Most of the improvement is attributed to *taigen-dome*. The worsened case is attributed mainly to the inability of the translation unit to parse the headline which was augmented with the CR result and became lengthy.

The upper part of Figure 5 shows an improved case. The underlines indicate the change of target words and the doubled underlines show the stylistic improvement. The target word of “CD” and “ceiling” improved from “C D (literal translation)” to “譲渡可能定期預金証書 (certificate of deposit)”¹⁰ and “天井 (*tenjou*, ceiling of a room)” to “上限 (*jougen*, upper limit),”, respectively. The translation style is more like a Japanese headline in terms of “、 (*kuten*)” and *taigen-dome* at the end of the sentence. Another improved case is shown in Figure 2, which is a citation from the result of experiment.

On the other hand, most of the worsened cases are attributed to the errors of the CR unit. The lower part of Figure 5 shows a worsened case where the wrong coreference between “RJR Nabisco” in the headline and “a shareholder of RJR Nabisco Holdings Corp” in the lead produced the wrong sentence, meaning “a shareholder sues a shareholder”.

Another remaining problem is that for about half of the test articles, the SAR unit could not retrieve any article from the corpus which deals with the same event as the input article. Moreover, no target words were applicable to half of the test articles for which the SAR unit retrieved some relevant articles and the TWE unit extracted some target words. It suggests that a size of 30,000 articles is too small to cover the variety of news event types, and also too small to cover the appearance of major distinct target words relevant to an event type.

6 Conclusion

This paper proposed a new approach to translating English news headlines. In this approach, the abbreviations, acronyms, and omissions in

⁸The system has been implemented on a SUN-Workstation Ultra60 450MHz in perl and partly in C where speed was important.

⁹The average word count and the byte size per article are approx. 120 and 773, respectively.

¹⁰This target word is found by coreference resolution.

headline: Shanghai New Asia to list 25 million A shares

lead: SHANGHAI, Oct 8 (Reuters) - Hotel operator Shanghai New Asia (Group) will list 25 million domestic A shares on Friday on the Shanghai stock exchange, a company statement said on Tuesday.

(original trans.) 上海新しいアジアは2500万A株をリストする。
 Shanghai - new - Asia to make an ordered array of items
 normal sentence end “する。”(suru)

(headline trans.)
 ホテル経営者上海ニュー・アジア、2500万A株を上場
 Hotel operator Shanghai New Asia to list stocks on the stock market
 taigen-dome(absence of “する。”)

Figure 2: comparison of the translation result

similarity	headline Japanese headline(translated by hand)
0.76	Tibet Jinzhu to list 13.5 mln A shares on Nov 8 Tibet Jinzhu (中国)、11月8日に上海証券所にA株1350万株を上場へ
0.72	China Inner Mongolia Hongfeng to list A shares Inner Mongolia Hongfeng (中国)、深セン証券取引所にA株を上場へ
0.67	Gintian eyes Singapore secondary listing Gintian (中)、シンガポール上場を計画
0.64	First Tibet industrial company to list A shares チベットの鉱工業関連会社、初の上海A株上場へ=証券専門紙
0.63	Seoul's new share listings down 35.6 pct in '95 ソウル株式市場、95年の新規上場35.6%減=K S E
0.63	Guangshen Rail issues over-allotment shares Guangshen Rail (中国)、香港株式市場で新株を追加発行へ
0.62	China distillery Gujing to list 60 mln B shares Anhui Gujing Distillery、12日にB株6000万株を深センに上場
0.61	Xinjiang Huitong to issue 12.5 mln A shares 中国新租の発電所建設会社、A株1250万株を発行へ
0.61	Padaeng to issue 122 million new shares パダエン(タイ)が1億2200万株を新規発行へ

Figure 3: Similar Articles retrieved from a Parallel Corpus

the headline are complemented by coreference resolution between the headline and the lead. Also, target words are selected by referring to the translation of similar news articles retrieved from a parallel corpus.

The experiment showed that this approach is effective. A major contribution is attributed to coreference resolution. Also, the style of translation improved because of the headline-specific translation rules. The target words extracted from the parallel corpus proved to be precise, but among them, only a few were applicable in

the translation.

The coreference resolution(CR) unit has been ported into our commercial English to Japanese MT system and is now a part of its Web page translation function.

References

- Claire Grover Andrei Mikheev and Marc Moens. '98. Description of the LTG system used for MUC-7. In *Proc. 7th Message Understanding Conference(MUC-7)*, Morgan Kaufmann, San Mateo, CA.
- Nigel Collier, Akira Kumano, and Hideki Hirakawa. 98. English-Japanese news article alignment from the In-

source word	original(default) target word extracted target word
issue(n)	問題 (“ <i>mondai</i> ”, a point of issue, the matter in question) 発行 (“ <i>hakkou</i> ”, publication, the act of issuing printed materials)
list(v)	リストする (“ <i>lisuto suru</i> ”, to make an ordered array of items) 上場する (“ <i>jōjō suru</i> ”, to list stocks on the stock market)
new(adj)	新しい (“ <i>atarashii</i> ”, new) 新規 (“ <i>shin-ki</i> ”, formal expression of ‘new’)
raise(v)	上げる (“ <i>ageru</i> ”, to raise) 増額する (“ <i>zougaku suru</i> ”, to increase sum of money)
statement(n)	ステートメント (“ <i>sutētomento</i> ”, literal translation of ‘statement’) 声明 (“ <i>seimei</i> ”, announcement)
total(n)	合計 (“ <i>goukei</i> ”, sum total) 総数 (“ <i>sousuu</i> ”, total number)

Figure 4: Extracted translation(target word)

headline: S. Korea to raise ceiling on banks’ CD issuance

(original) 韓国は銀行のCD配給で天井を調達する。

literal trans. ceiling of a room

(headline trans.)

韓国、銀行の譲渡可能定期預金証書の配給で上限を調達
certificate of deposit upper limit *taigen-dome*(absence of “する”)

headline: Shareholder sues RJR Nabisco

(original) 株主はRJRナビスコを訴える。

a shareholder RJR Nabisco sues

(headline trans.)

株主、RJRナビスコ・ホールディングズ社のA株主を訴える
a shareholder a shareholder of RJR Nabisco Holdings Corp sues

Figure 5: Example of improved and worsened translation

- ternet using MT. In *Japan Soc. for AI Annual Meeting*.
- Zheng Jie and Mao Yuhang. 2000. A Word Sense Disambiguation Method Using Bilingual Corpus. In *second Int. Conference on Language Resources and Evaluation(LREC2000)*.
- Hiroyuki Kaji and Yasutsugu Morimoto. 2002. Unsupervised Word Sense Disambiguation Using Bilingual Comparable Corpora. In *Proc. Int. Conference on Computational Linguistics(COLING ’02)*.
- George R. Krupka and Kevin Hausman. ’98. IsoQuest INC.: Description of the NetOwl Extractor System as Used for MUC-7. In *Proc. 7th Message Understanding Conference(MUC-7)*, Morgan Kaufmann, San Mateo, CA.
- Vincent Ng and Claire Cardie. 02a. Identifying Anaphoric and Non-Anaphoric Noun Phrases to Improve Coreference Resolution. In *Proc. Int. Conference on Computational Linguistics(COLING ’02)*.
- Vincent Ng and Claire Cardie. 02b. Improving Machine Learning Approaches to Coreference Resolution. In *Proc. 40th Annual Meeting of the Association for Computational Linguistics(ACL 02)*, pages 103–111.
- Ichiro Ooi and Tatsuhiko Tsunoda. 96. Disambiguation of Nouns in Headlines of English Newspapers Based on Lexical Cohesion. In *Technical report of IEICE NLC 96-2(in Japanese)*, pages 9–14.
- Satoshi Shirai, Yoshifumi Ooyama, and Yohsataka Nakao. 97. Characteristics of English Newspaper Article Headlines. In *Proc. 54th Annual Convention, 4B-1, Inf. Proc. Soc. Japan(in Japanese)*.
- Yamato Takahashi, Satoshi Shirai, and Susumu Fujinami. 96. Automatically Aligning Japanese & English Newspaper Articles. In *IPSJ SIG Notes, NL-114-9(in Japanese)*, pages 55–62.
- Takehiko Yoshimi and Ichiko Sata. 99. Improvement of Translation Quality of English Newspaper Headlines by Automatic Preediting. In *MT Summit VII*, pages 496–500.
- Yumiko Yoshimura, Satoshi Kinoshita, and Miwako Shimazu. 97. Processing of Proper Nouns and Use of Estimated Subject Area for Web Page Translation. In *Proc. 7th International Conference on Theoretical and Methodological Issues in Machine Translation(TMI)*, pages 10–18.