

# Transforming a Sentence End into News Headline Style

Satoshi Ikeda and Kazuhide Yamamoto

Dept. of Electrical Engineering, Nagaoka University of Technology  
iked@nlp.nagaokaut.ac.jp, yamamoto@fw.ipsj.or.jp

## Abstract

News on electrical bulletin boards consist of high density expressions. Many sentences end with unique expressions that consist of nouns and case particles. This paper focuses on expressions used at the end of sentences and attempts to summarize them by forming noun or case particle endings. We summarize the news sentence through pattern matching approach. Our evaluation illustrates that the summarizer reduces 2.50 characters per sentence on average; the reduction ratio is 6%. We also show that people perceive the correct meanings of the summarized sentences with 95% accuracy.

## 1 Introduction

Electrical bulletin board displays the latest news headlines which each newspaper office announced. News headlines are shorter than newspapers' news with laconicism because they are summarized to transfer in limited space.

One of a characteristics of Japanese news headlines can be seen at sentence ends (Exp.1).

Exp.1) 拉致疑惑は与党の動向見極め 判断。  
(Countermeasure for alleged abduction is judged after the movement of administration party.)

Although the end of sentence in Exp.1 is omitted, we have no difficulty to understand the meaning. We unconsciously complete the sentence by guessing what is omitted without a mistake. Without unnecessary ends, these type of sentences are short and nonredundant.

The final purpose of this work is to transform news sentences into a news headline style. In Japanese, sentences end with nouns or case particles are grammatically incorrect, however, this

kind of expressions are shorter than grammatically perfect sentences, and hence often used to meet the limited length. We believe many sentences have semantically redundant expressions in the end, which needs to be focused in summarization.

In this paper we present a list of deletable expressions at Japanese sentence ends. We have to carefully investigate which sentence ends are deletable, and how to change into the headline style. We present the concrete expressions of deletion with examples and illustrate effects of deletion.

## 2 Related Works

As the most similar work to ours, Sato et al.[6] tries to extract paraphrasing patterns of sentence end by preparing a lot of alignment pairs between news sentences and their headline versions. They compare the sentences from the ends and obtain many correspondences between the two. However, they have no proposals on how to use these one-to-N correspondences, i.e., the way to select one from many candidates. Our approach is to obtain many transformation patterns as well, but we do not use aligned corpus; we use a large collection of news headlines instead and find patterns by our thorough observation.

Wakao et al.[7] compares newscast and corresponding subtitle expressions to investigate the differences of them. One of the observation targets is sentence end, and they have shown us some typical patterns of conversion into a short news. This enumerates phraseologies which are able to be cut down and investigates the frequency of use. In news subtitles nouns or case particles are used at the sentence end. This work is drew upon literature [7] while we investigated in our own right. We shaded light on the phraseologies which do not exist literature [7] such as 「自首したと見られる 自首か (There seems to surrender.)」. We examined the phraseologies which

are disposed by machine. Fukushima et al. [1] cut off the unnecessary part from literature [7].

There are investigations to summarize text by confining the number of characters [2,3,5]. Ishizako et al.[2] cut off areas of overlap. Ohmori et al.[5] and Mikami et al.[3] summarized text altogether, but these investigations do not focus on sentence ends.

### 3 News Headlines and Their Sentence Ends

There is an email service that delivers Japanese news headlines three times a day on weekdays. That is Nikkei news mail(1) provided by NIKKEI-goo. We have been collecting them since December 1999. Table1 shows the statistics we have obtained.

Table 1: Statistical datum which are collected

number of mails	3365
number of stories	21127
number of sentences	40374

News headlines are more distinctive than news stories in sentence end. Therefore, we investigated part of speech on both news headlines and newspaper(Nihon Keizai Shimbun(2)). Table2 shows the comparison.

Table 2: Occurrence ratio of POS in sentence end

POS	occurrence ratio[%]	
	newspaper	headlines
noun (verbal noun)	23.70 (5.00)	55.92 (39.90)
verb	28.66	15.91
adjective	1.80	0.19
adverb	0.20	0.22
particle (case particles)	1.56 (0.34)	8.83 (6.41)
auxiliary verb	38.59	18.52
symbol	5.42	0.40

In the newspaper, declinable words are responsible for the majority of sentence ends. In news headlines, there are in fact many verbal nouns in sentence ends.

Japanese words are classified broadly into two types; one derived from China and another origi-

nated in Japan. News headlines contain the former more than the latter because words from China carry more information in fewer characters. We investigated news headlines and news on a paper which contained words of both Chinese and Japanese origins. The result is shown in Table3. In fact, news headlines preferably use the words of Chinese origin about three times as much as that of Japanese origin.

Table 3: Ratios of Chinese and Japanese origin words in (a) newspaper and (b) headlines.

Japanese	Chinese	ratios [%]		a/b
		(a)	(b)	
見つかる	発見 (to be found)	1.059	2.658	0.398
決める	決定 (to decide)	0.622	2.184	0.285
選ぶ	選出 (to elect)	0.210	2.643	0.079
分かる	判明 (to find out)	0.181	2.875	0.063
命じる	命令 (to order)	1.132	3.841	0.295
述べる	発言 (to say)	0.456	0.181	2.493
調べる	調査 (to investigate)	6.284	53.333	0.118
total		2.712	7.271	0.373

We can imagine that a short phraseology is preferably used when the phraseology has the same information. We estimate that the news headlines are high density phraseology than newspaper.

### 4 Method of Summarization

In order to transform a sentence end into a shorter one, we have conducted three kinds of procedures:

- (1) Deletion of target words at sentence end
- (2) Deletion with minor transformation after the target words
- (3) Transformation of sentence end

More precisely, we have proposed conducting the following 10 procedures for transforming Japanese sentence ends into a news headline style:

1. Cut off dictum and honorific phraseology (1)
2. Cut off 「を示す (wo shimesu:show)」 (1)
3. Change verbal noun(2)
4. Cut off 「なる (naru)」 (2)
5. Cut off the part which follows 「明らかに (akirakani)」 (2)
6. Change words of Japanese origin(2)
7. Cut off 「てしまう (teshimau)」 (1)
8. Cut off 「立つ (tatu)」 (2)
9. Transform phraseology indicated the action in the future (2)

#### 10. Change to compound noun (3)

We summarized in this order, and process 3.~  
9. can be switched.

### 4.1 Cut Off Dictum and Honorific Phraseology

Phraseologies shown below are dictum or honorific phraseology. These phraseology in sentence end is cut off because these are not necessary to understand the meaning.

- dictum phraseology: 「だった (datta)」 「である (dearu)」 「だ (da)」
- honorific phraseology: 「ます (masu)」 「です (desu)」

### 4.2 Cut Off 「を示す (wo simesu:show)」

When a sentence end is 「を示す (wo shimesu)」 or 「を示した (wo shimeshita)」, this phraseology is cut off because 「示す (shimesu)」 has little meaning in that sentence. The main verb of the sentence is the verbal noun before 「を示す (wo shimesu)」.

### 4.3 Change Verbal Nouns

The expression after the verbal noun closest to the main verb of the sentence is deleted. In Japanese, we put a word 「する (suru)」 after a verbal noun to make a verb, but in the summary it can be deleted since we can still understand the usage.

When a self-sufficient word exists following a verbal noun, we do not dispose this.

**Step 1** The part following 「する (suru)」 is cut. Nominalized verbal noun to cut 「する (suru)」 is the verbal noun in this arrangement.

**Step 2** When the cut part contains an estimation phraseology 「みられる (mirareru)」 or 「だろう (daou)」, tack on 「か (ka)」 and finish.

Exp.2) 逃走資金に困って自首 したとみられる。  
逃走資金に困って自首 か。

(He seemed to surrender in trouble with escape fund.)

**Step 3** When the cut part contains a contradiction phraseology 「ない (nai)」 or 「ぬ (nu)」, tack on 「せず (sezu)」 at the sentence end and finish. When this part concurrently contains a passive phraseology 「れる (reru)」, tack on 「されず (sarezu)」 and finish.

**Step 4** When a sentence end is 「noun + を (wo) + verbal noun」, 「を (wo)」 is cut to become a compound noun 「noun + verbal noun」.

Exp.3) 今月から各駅と車内のつり広告に  
詰め将棋を掲載する。  
今月から、各駅と車内のつり広告に  
詰め将棋掲載。

(Starting this month, Japanese chess problems are seen in ads of each station and in trains.)

**Step 5** When a sentence end is 「particle<sub>1</sub> + noun + すること (surukoto) + particle<sub>2</sub> + noun」, 「すること (surukoto)」 is cut. If the particle<sub>1</sub> is 「を (wo)」 or 「か (ka)」, this particle changes to 「の (no)」.

If the cut part contains 「初めて (hajimete:first)」, procedures from Step 2 is different as follows.

**Step 2** When the cut part contains 「するのは (surunoha)」 or 「したのは (shitano)」, 「初めて (hajimete:first)」 is tacked on before verbal noun. When the part of cut contains 「みられる (mirareru)」, tack on 「か (ka)」 in sentence end and finish.

**Step 3** When the cut part contains 「して (shite)」, 「後初 (go hatsu)」 is tacked on in the sentence end. When the term just before noun is particle 「か (ka)」, this particle 「か (ka)」 changes into particle 「の (no)」.

Exp.4) 会見に応じたのはカルマパ17世が中国から  
出国 して初めて。  
会見に応じたのはカルマパ17世が中国から  
出国 後初。

(He first acceded the interview since Karmapa Seventeenth left China.)

**Step 4 .1** When a verbal noun is 「発言 (hat-sugen:delivery)」 or 「言及 (genkyuu:citation)」, 「初めて (hajimete:first)」 is tacked on before the verbal noun.

Exp.5) ロシア軍幹部が撤退に 言及したのは初めて。  
ロシア軍幹部が撤退に 初めて言及。

(Russian troop's cadre first adverted to retreat.)

**Step 4.2** When a verbal noun is not 「発言 (hat-sugen:delivery)」 or 「言及 (genkyuu:citation)」, the term before the verbal noun is checked. The sentence end is processed the following.

- particle 「の (no)」, 「が (ga)」 + verbal noun  
particle 「の (no)」 + verbal noun + 「は初 (hatsu)」
- particle 「を (wo)」, 「も (mo)」 + verbal noun  
particle 「を (wo)」, 「も (mo)」 + verbal noun
- otherwise  
~ verbal noun ~ の + verbal noun + 「は初 (hatsu)」

**Step 5** When the cut part contains 「みられる (mirareru)」, 「か (ka)」 is tacked on in the sentence end.

#### 4.4 Cut Off 「なる (naru)」

When 「particle + なる (naru)」 exists in a sentence, this part and the following are cut off. When a self-sufficient word exists in the cut part, the meaning changes or we do not understand the meaning.

Therefore, when a self-sufficient word exists 「particle + なる (naru)」 following, the sentence is not disposed this arrangement.

The 「particle + なる (naru)」 and the following are cut off. When the particle is 「に (ni)」 or 「と (to)」, 「に (ni)」 is tacked on in the sentence end.

Exp.6) 総選挙投票3カ月半後の合意で、ぎりぎりの  
選択 となった。  
総選挙投票3カ月半後の合意で、ぎりぎりの  
選択 に。

(The accord became the bare adoption because this arranged after three and half months of general election ballot )

When the cut part contains a contradiction phraseology 「ない (nai)」 or 「ぬ (nu)」, 「ならず (narazu)」 is tacked on in the sentence end.

Exp.7) 火薬が湿っていたのか、ほとんどが起爆剤  
にならなかった。  
火薬が湿っていたのか、ほとんどが起爆剤  
にならず。

(Almost all detonating agents did not work because they seemed to wet)

#### 4.5 Cut Off the Part After 「明らかに」

When 「明らかに (akirakani:out of doubt)」 exists in a sentence, the part which follows 「明らかに (akirakani)」 is cut off. When the cut part contains a self-sufficient word, the meaning changes or we do not understand the meaning.

Then, when a self-sufficient word exists in the sentence, the sentence is not disposed this arrangement.

**Step 1** The part which follows 「明らかに (akirakani)」 is cut off.

**Step 2** Research the part of cut and dispose the cut part.

- Contradiction phraseology 「ない (nai)」 or 「ぬ (nu)」 and passive phraseology 「れる (reru)」 exist.  
「されず (sarezu)」 is tacked on in the sentence end.

- The contradiction phraseology 「ない (nai)」 or 「ぬ (nu)」 exists.  
「せず (sezu)」 is tacked on in the sentence end.

Exp.8) 特別損失額は明らかに していない。  
特別損失額は明らかに せず。(The amount of loss is not announced.)

**Step 3** When 「することを (surukoto wo)」 exists before 「明らかに (akirakani)」, 「することを (surukoto wo)」 is cut off. When the part before the cut is 「particle 「に (ni)」 + verbal noun」, 「に (ni)」 is changed to 「へ (e)」. When the part before the cut part is 「particle 「を (wo)」 + verbal noun」, 「を (wo)」 is changed to 「の (no)」.

#### 4.6 Change Words of Japanese Origin

When a Japanese origin word by Table3 exists in a sentence, the part before it is cut off. Then the Japanese origin word is replaced by Chinese one.

When a self-sufficient word exists following Japanese origin word, the sentence is not disposed of this arrangement. We changed the word which shows Table3.

**Step 1** Japanese origin word and following are cut off.

**Step 2** When sentence end is 「することを (surukoto wo)」, cut off 「すること (surukoto:doing)」, tack on the correspondent Chinese origin word, and finish the arrangement.

Exp.9) 「災害広域支援マニュアル」の作成に着手 することを決めた。  
「災害広域支援マニュアル」の作成に着手  
を決定。

(They have decided to start making an 'instruction book on extensive assistance for disaster'.)

**Step 3** When a sentence condition is followed, the sentence is disposed.

- A sentence end is a particle 「が (ga)」 and Japanese origin word is 「分かる (wakaru:understand)」  
tack on 「判明 (hanmei:understand)」 and finish.
- A sentence end is a particle 「が (ga)」 and Japanese origin word is not 「調べる (siraberu:census)」  
The particle 「が (ga)」 is changed to a particle 「を (wo)」.
- A sentence end is 「が (ga) + noun + で (de)」 「の (no) + noun + を (wo)」
- A sentence end is a particle 「は (ha)」 and Japanese origin word is 「分かる (wakaru:understand)」  
Get the former sentence back again and finish.
- Japanese origin word is 「調べる (siraberu:census)」 and the cut part contains 「している (siteiru)」

tack on 「調査中 (tyousa tyu:under survey)」 at the sentence end and finish.

**Step 4** Chinese origin word which corresponds Japanese origin word tacked on the sentence end.

Exp.10) 変造硬貨計 359 枚 が見つかった。  
変造硬貨計 359 枚 を発見。

(The total of 359 counterfeit coins were found.)

#### 4.7 Cut Off 「てしまう (teshimau)」

When a sentence contains 「てしまう (teshimau)」, we feel that the sentence is negative and 「てしまう (teshimau)」 is not necessary to understand the meaning of the sentence. Thus we cut off 「てしまう (teshimau)」 in the headline. This arrangement is used not only the sentence ends but middle of the sentence. When the term after the cut part is 「ば (ba)」, we do not dispose it. When the sentence end is 「てしまう (teshimau)」, change the term before 「てしまう (teshimau)」 to primitive form and finish.

When 「てしまう (teshimau)」 exists without the sentence end, 「てしまう (teshimau)」 and the character before this phraseology is cut off.

#### 4.8 Cut off 「立つ (tatsu)」

When a sentence contains 「立つ (tatsu)」, 「立つ (tatsu)」, the part following it is cut off. When the following part contains the self-sufficient word, the meaning changes or we do not understand the meaning.

Therefore, when a self-sufficient word exists in the following part, the sentence is not disposed this arrangement. When 「立つ (tatsu)」 is a part of idiom, the sentence is not disposed of this arrangement.

**Step 1** 「立つ (tatsu)」 and the following part are cut off.

Exp.11) 「トップボーイ」はTVゲームの専門小売店の  
頂点に 立つ。  
「トップボーイ」はTVゲームの専門小売店の  
頂点に。

(‘Top boy’ is acme in TV game retail business)

**Step 2** When a contradiction phraseology 「ない (nai)」 or 「ぬ (nu)」 exists in the cut part, 「立たず (tatazu)」 is tacked on at the sentence end.

#### 4.9 Phraseology of Words Implying Future

When a phraseology which indicate the action in the future such as 「計画 (keikaku:attempt)」

or 「予定 (yotei:plan)」 exists in the sentence, the phraseology can be changed to 「へ (he)」 in Japanese. Therefore, the terms listed below are the phraseology of indicated the action in the future. When 「する (suru) + this phraseology」 exists in the sentence, this part and following are changed to 「へ (he)」.

「予定 (yotei:plan)」 「計画 (keikaku:attempt)」 「方針 (houshin:policy)」 「方向 (houkou:future direction)」

When 「する (suru) + this phraseology」 exists in a sentence and the following contains a contradiction phraseology 「ない (nai)」 or 「ぬ (nu)」, the sentence is not disposed of this arrangement. When the following contains the 「という (toiu)」 or 「、」, the sentence is not disposed of this arrangement.

「する (suru) + this phraseology」 and following are cut off. when the sentence end is particle, the particle is cut off. 「へ (he)」 is tacked on the sentence end.

#### 4.10 Change to a Compound Noun

When a sentence end is 「noun + particle + verbal noun」 after the above arrangements, the particle cut off to become a compound noun. When the noun is neither pronoun, person name, unique noun nor postfix for Chasen(3), this arrangement is not disposed. When the particle is 「から (kara)」, 「で (de)」 or 「も (mo)」, this arrangement is not disposed.

We make a compound noun dictionary for The Mainichi Newspapers(4) to check the adequacy of compound nouns. When 「noun particle に (ni) verbal noun」 and the dictionary contains 「noun + verbal noun」 which is cut of 「に」, 「noun + particle に (ni) + verbal noun」 is changed to 「noun + verbal noun」. When the particle is not 「に (ni)」, 「noun + particle + verbal noun」 is changed to 「noun + verbal noun」.

Exp.12) 3階の焼け跡から男性の遺体 が見つかった。

3階の焼け跡から男性の 遺体を発見。

3階の焼け跡から男性の 遺体発見。

(A man’s body was found on the third floor of burned-out site.)

## 5 Experiments

We implemented the proposed technique with Perl programming language to measure the ade-

quacy of proposed technique. We summary with this program. Then input sentence are all sentences seen in the newspaper corpus. The number of input sentences is 232,038, and 73,512 outputs are somehow summarized in our method.

### 5.1 Summarization Ratio

We calculated a sentence ratio and number of reduced characters in a sentence. This result of experiment is shown in Table4. The method of Table4 shows the section number. This Table4 shows the result which used the only one method. The summarization ratio is 94%. In fact, this method is reduced the 6% about one sentence.

Table 4: Summarization ratio

process	4.1	4.2	4.3	4.4	4.5
# sentence	16825	1313	37995	7510	199
summ. ratio	0.94	0.94	0.94	0.93	0.90
# reduced char.	1.60	4.00	2.56	3.12	5.41
process	4.6	4.7	4.8	4.9	total
# sentence	7194	600	197	848	72681
summ. ratio	0.96	0.89	0.92	0.87	0.94
# reduced char.	2.20	3.93	3.28	6.57	2.45

### 5.2 Subjective Evaluation

We also evaluated the proposed technique by human judgment. We picked up 1,000 sentences at random from summary sentences, and three examinees individually accounted them. The sentences are measured by majority decision. Assessment criterion is: (1) same meaning without context, and (2) low unnaturalness. The result is shown in Table5. The numbers in the table denote the section numbers explaining the process of transformation.

Table 5: Correctness of each process

method	4.1	4.2	4.3	4.4	4.5
# sentence	231	19	492	107	9
# correct	205	18	481	106	8
ratio	0.89	0.95	0.98	0.99	0.89
method	4.6	4.7	4.8	4.9	total
# sentence	116	21	3	13	1000
# correct	113	17	3	12	952
ratio	0.97	0.81	1	0.92	0.95

We have also computed the influence of personal difference. In this kind of subjective evalua-

tion different person may answer difference judgment. We have evaluated our results in three criteria: (1) at least one said correct, (2) at least two said correct, and (3) all three said correct. This result is shown in Table6. The Table illustrates that correctness is more than 90% in all cases.

Table 6: Correctness changes by personal differences.

	$\geq 1$	$\geq 2$	$= 3$
correctness	0.98	0.95	0.91

### 5.3 Comparison to the Human Summaries

We compare summaries of the proposed method and by the human. We picked up 100 sentences in summary sentences at random. One examinee summarized the original sentences which corresponded the pick up the summary sentences. We computed the summarization ratio about these sentences. The result is shown in Table7.

Table 7: Comparison of summaries by proposed technique and manual summary

	machine	human
# sentence	72727	100
summ. ratio	0.94	0.92
# reduced characters	2.45	3.87

Although the sentence ratio of machine summary is close to the manual summary's one, number of reduced characters are approximately one character different. This indicates that human try to change many parts of sentence according to the change of the sentence end, while the machine does not consider such influence. Change of sentence end often requires transforming the whole syntax structure, such as change of aspect or form. We need more investigations on this issue.

## 6 Discussions

### 6.1 Discussion of Erroneous Summaries

In this section we describe some erroneous summaries by our method and discuss the reasons.

Exp.13) 顔はその人の年輪みたいなもので、喜怒哀楽の表情を 15.5mm 積み重ね、人柄を示す。

\*<sup>1</sup>顔はその人の年輪みたいなもので、喜怒哀楽の表情を積み重ね、人柄。

(The face show the character like an annual ring.)

Exp.13 is error example in arrangement 'cut off the 「を示す (wo shimesu:show)」'. When the term before 「を示す (wo shimesu)」 is the noun, the sentence does not have main verb. The main verb which does not exist in the sentence is not right in Japanese. when the term before 「を示す (wo shimesu)」 is noun, this arrangement does not disposed. This kind of error is covered. But when the noun is 「考え (kangae:concept)」, 「意向 (ikou:disposition)」 or 「見通し (mitooshi:forecast)」, this arrangement is correct.

Exp.14) 利用者に過度の使用を警告することを決めた。

\*利用者に過度の使用 を 警告を決定。

(It is decided to caution the overuse to user.)

Exp.14 is the error example 'change the word of Japanese origin. When the cut off 「すること (surukoto:doing)」, the modification relation is changed. Therefore, the modification relation is a wrong one. When the particle 「を (wo)」 is changed to particle 「の (no)」, this kind of error is covered(Exp.15).

Exp.15) 利用者に過度の使用を警告することを決めた。  
利用者に過度の使用 の 警告を決定。

Exp.16) 母親を殺してしまおうと思っていた。

\*母親を 殺しう と思っていた。

(He thinks that his mother killed.)

Exp.16 is an error example in 'cut off 「てしま (teshimau)」'. When 「てしま (tesimau)」 is cut off, it is not congruent inflected forms of 「てしま (teshimau)」 and the verb. When 「てしま (teshimau)」 is cut off, the inflected forms must be congruent.

## 6.2 Verbalness/Nominalness of Verbal Noun

The sentence end is 「～は初 (ha hatsu:first)」 in Section4.3. There are a big differences by humans in degree of accepting this expression. We thus change expression 「～は初 (ha hatsu)」 into 「～初めて (hajimete:first)～」. The example before changed is shown in Exp.17.

Exp.17) プーチン大統領がアラブ国家の指導者と

会談するのは初めて。

プーチン大統領がアラブ国家の指導者との

会談は初。

<sup>1</sup>symbol '\*' indicates that the sentence is wrong.

(It is the first time that President Putin has a talk to the captain of Arab Crown)

Some people feel unnatural or wrong in this example. But when the original sentences do not have 「初めて (hajimete:first)」, the summary sentences are correct. The example is shown Exp.18 without 「初めて (hajimete:first)」.

Exp.18) プーチン大統領がアラブ国家の指導者と  
会談する。

プーチン大統領がアラブ国家の指導者と 会談。

This example gives us no unnaturalness. We think the verbal noun affect this. The verbal noun represents that indicates the kind. The verbal operation of verbal noun is varied by humans.

We think concretely about 「会談 (kaidan:meeting)」 of Exp.17and Exp.18. First, we think that 「会談 (kaidan:meeting)」 is complemented the verbal noun 「会談する (kaidan suru:have a talk)」. The predicate is generally at sentence end in Japanese. When the predicate does not exist in a sentence, it is inclinable in human thought that sentence end term is predicate. The other hand, we think that 「会談 (kaidan:colloquy)」 is nominal or verbal operation in Exp.18 because 「会談 (kaidan:meeing)」 is not sentence end. then when 「会談 (kaidan:meeting)」 is nominal, human have unnaturalness. And when 「会談 (kaidan:meeting)」 is verbal, human do not feel unnatural.

We cite the error summary which sentence end is noun other than verbal noun in this paper but the verbal operation of noun is pertained in these sentences. And the noun of operation verbal is 「考え (kangae:concept)」 other than verbal noun.

## 6.3 Comparison of Machine and Manual Summaries

We examine the machine and manual summaries. Although many sentences are not much different, some sentences have big differences for summarization. One example is shown as follows, original sentence, its machine summary and its manual summary respectively.

Exp.19) カラー写真を使ったグラフ面なども あります。

カラー写真を使ったグラフ面なども ある。

カラー写真を使ったグラフ面なども。

(There is graph used the color picture.)

Exp.19 is cut off the honorific phraseology but the manual summary is cut off 「ある (aru)」 too.

This is shown that 「ある (aru)」 is dictum phraseology. And the sentence end is 「も (mo)」. This is often seen in the news headline. But the proposed technique do not deal with them.

#### 6.4 Summarization Failure

We examine the sentences which are not summarized by the method. We picked up the 200 sentences at random and examine whether or not it should be summarized. This results is that 9 sentences are missing. The example is shown below with the supposed summary.

Exp.20) 焼け跡から池本さんが遺体で発見  
された。

焼け跡から池本さんが遺体で発見。

(Mr. Ikemoto's blob was found from the burned-out site.)

Exp.20 is not summarized. The reason of this error is caused by an error of the morphological analysis.

#### 7 Conclusion

In order to generate short and smart style seen in news headlines, this paper presents a method of transforming Japanese sentence end expressions into short style. Our observation reveals that the end of sentence in the headlines are either nouns or case particles in many sentences, we thus attempt to summarize them as short as possible. We have implemented the approach and evaluated in summarization ratio and their correctness. The results illustrates that the reduction ratio is 6% against overall sentence length, and the sentence is expected to be cut off 2.50 characters per sentence. The length of automatic shortening is approximately the same as manual summarization. We also confirmed that 95% of the summaries were judged to be correct.

#### Acknowledgment

This work was supported in part by MEXT Grants-in-Aid for Young Scientists (B) 16700134, and for Scientific Research (A) 16200009, Japan.

#### Tools and language resources

- (1) Nikkei news mail, NIKKEI-goo,  
<http://nikkeimail.goo.ne.jp/>
- (2) Nihon Keizai Shimbun Newspaper Corpus,  
year 2000, Nihon Keizai Shimbun, Inc.
- (3) Chasen, Ver.2.3.3, Matsumoto Lab, Nara Institute of Science and Technology.

<http://chasen.naist.jp/hiki/ChaSen/>

- (4) The Mainichi Newspapers Corpus, year 2000,  
Mainichi Newspaper Co., Ltd.

#### References

- [1] Takahiro Fukushima, Terumasa Ehara and Katsuhiko Shirai. 1999. Regulation for Reducing Number of Characters for Sentence Simplification, Proceedings of The Fifth Annual Meeting of The Association for Natural Language Processing, pp.221–224. (in Japanese)
- [2] Yuko Ishizako, Akira Kataoka, Shigeru Masuyama and Seiichi Nakagawa. 1999. Summarization by Reducing Overlaps and Its Application to TV News Texts. IPSJ SIG Technical Reports 99-NL-133(7), pages 45–52. Information Processing Society of Japan. (in Japanese)
- [3] Makoto Mikami, Shigeru Masuyama and Seiichi Nakagawa. 1999. A Summarization Method by Reducing Redundancy of Each Sentence for Making Captions of Newscasting. Journal of Natural Language Processing Vol.6, No.6, pp.65–81. (in Japanese)
- [4] Kiyonori Ohtake and Kazuhide Yamamoto. 2001. Paraphrasing Honorifics. Proc. of NLPRS2001 Post-Conference Workshop on Automatic Paraphrasing: Theories and Applications, pp. 13–20.
- [5] Takefumi Oomori, Hidetaka Masuda and Hiroshi Nakagawa. 2003. Web News Articles Summarization and its Evaluation using Articles for Mobile Terminals, IPSJ SIG Technical Reports 2003-NL-153(1). pages 1–8. Information Processing Society of Japan. (in Japanese)
- [6] Dai Sato, Moritaka Iwakoshi, Hidetaka Masuda and Hiroshi Nakagawa. 2004. Extraction of Paraphrasing Patterns from Aligned Corpora of Web and Mobile Terminal News Articles. IPSJ SIG Technical Reports 2004-NL-159(27). pages 193–200. Information Processing Society of Japan. (in Japanese)
- [7] Takahiro Wakao, Terumasa Ehara and Katsuhiko Shirai. 1997. Summarization Methods Used for Caption in TV News Programs, IPSJ SIG Technical Reports 97-NL-122(13). pages 83–89. Information Processing Society of Japan. (in Japanese)