Unsupervised Knowledge Acquisition about the Deletion Possibility of Adnominal Verb Phrases

Hiroyuki SAKAI and **Shigeru MASUYAMA**
Department of Knowledge-based Information Engineering,
Toyohashi University of Technology,
Toyohashi 441-8580, Japan
sakai@smlab.tutkie.tut.ac.jp, masuyama@tutkie.tut.ac.jp

Abstract

This paper addresses an unsupervised learning method which acquires knowledge about the deletion possibility of adnominal verb phrases from corpus. Actually, our method deletes an adnominal verb phrase with which the noun modified by the phrase can be associated easily. Experimental results show that our method is able to delete adnominal verb phrases appropriately, as the precision and the recall attain 79.3% and 72.7%, respectively.

1 Introduction

Recent rapid progress of computer and communication technologies enabled us to access enormous amount of machine-readable information easily. This, however, has caused so called the information overload problem. Under these circumstances, the necessity for automatic summarization has been increasing and has been intensively studied recently (see e.g,(Mani and T.Maybury, 1999)). No all-purpose automatic summarization method for summarizing any type of documents appropriately exists, thus, in many cases, we must combine several summarization methods as components when we develop an automatic summarization system. Deletion of some unnecessary parts from a sentence is among such important summarization methods(Yamamoto et al., 1995)(Jing and McKeown, 2000).

In this paper, we propose an unsupervised learning method which acquires knowledge about the deletion possibility of adnominal verb phrases in order to summarize a document by deleting unimportant segments from a sentence in the document. As a method of summarization which deletes unimportant segments from a sentence, Ohtake et al.(Ohtake et al., 2001) proposes a method which deletes one of double modification segments modifying the same noun by using manually constructed rules. Knight et al.(Knight and Marcu, 2000) and Jing(Jing, 2000) and Takeuchi et al. (Takeuchi and Matsumoto, 2001) propose methods which extract rules for reducing a sentence from aligned corpus of human-made summaries and their original manuscripts. As for such application of supervised learning using aligned corpus, Katoh et al.(Katoh and Uratani, 1999) proposes a method which acquires knowledge about paraphrasing by using an aligned corpus of captions and original manuscripts which correspond to the captions. However, these previously proposed methods have the following drawbacks for practical use.

- To make a complete list of rules manually in order to delete unimportant segments from a sentence is a hard task.
- Aligned corpus between original manuscripts and summaries is useful. But it is not necessarily available. Moreover, constructing summaries manually for obtaining such aligned corpus is a time-consuming and costly task.

By these reasons, we propose an unsupervised learning method which acquires knowledge from general corpus (e.g., news paper articles), about the deletion possibility of adnominal verb phrases to summarize by deleting unimportant segments from a sentence. We define, in this paper, adnominal verb phrases as phrases which modify a noun and include a verb modifying the noun. An example is shown as follows.

Example 1: *Oobutai de engisuru sutaa* (大舞台で演技するスター: A star acting on a big stage),
Adnominal verb phrases: *Oobutai de engisuru* (大舞台で演技する: Acting on a big stage),

Including a verb: *engisuru* (演技する: act),
Modified a noun: *sutaa* (スター: star).

□

The method proposed by Ohtake et al.(Ohtake et al., 2001) deletes one of double adnominal modification expressions which are part of noun phrases by using manually-constructed rules. The method can only apply to sentences having structures matching the rules. In contrast, our method deletes adnominal verb phrases having unimportant contents by using statistical information obtained from single corpus. Consequently, more flexibility in deletion of adnominal verb phrases is attained.

Our method extracts knowledge about deletion of adnominal verb phrases from general corpus, e.g., news paper articles provided as a machine-readable form, documents obtained from WWW which are easily available.

We introduce our method in Sec. 2 and its implementation and experiments for evaluation are illustrated in Secs. 3 and 4, respectively. We analyze the results of the experiments in Sec. 5. Sec. 6 concludes this paper.

2 Proposed method

Our method is based on an intuition that adnominal verb phrases with which the modified noun is easily associated can be deleted. Here, an adnominal verb phrase is a phrase which modifies a noun and includes a verb modifying the noun (see Example 1). Among such adnominal verb phrases we consider, in this method,

- those modifying nouns not frequently modified by adnominal verb phrases and modified by only a few kinds of verbs,
- those modifying nouns having important contents,
- those having important contents.

Adnominal verb phrases modifying nouns not frequently modified by verbs and modified by only a few kinds of verbs can be deleted, as it is easy to associate such a noun with adnominal verb phrases modifying the noun and then deletion does not cause serious information loss. Our method gets, for each noun, "frequency of modification of the noun by adnominal verb phrases", "variety of adnominal verb phrases modifying the noun" etc., from news paper corpus provided as a machine-readable form, and acquires knowledge for deleting adnominal verb phrases by using their statistical information. In order to reflect "frequency of modification of the noun by adnominal verb phrases" and "variety of adnominal verb phrases modifying the noun", the entropy based on a probability of verbs to modify noun is employed. And, adnominal verb phrases modifying nouns which have large entropy are deleted by our method.

Adnominal verb phrases modifying nouns having important contents can be deleted, as meaning of such a noun is understood only by the noun itself, and then deletion of modifier does not cause serious information loss. Our method deletes adnominal verb phrases modifying nouns having important contents. In order to measure the importance of the content of a noun, we use *idf* (Salton, 1988) method. And, the adnominal verb phrases modifying nouns with large *idf* value (the nouns have important contents) are deleted.

We regard an adnominal verb phrase has important contents when

- it has nouns having large *idf* value,
- it has nouns appearing in some sentence after its position in the documents.

On the other hand, the importance of an adnominal verb phrase is reduced when

- it has nouns appearing in some sentence before its position in the documents.

The adnominal verb phrases having important contents and having noun appearing after its position in the documents tend not to be deleted. On the other hand, the adnominal verb phrases having noun appearing before its position in the documents tend to be deleted.

2.1 Algorithm to delete adnominal verb phrases

Based on the above observations, our method of deleting adnominal verb phrases is defined as follows:

Algorithm to delete adnominal verb phrases

Step 1: Weight value $W(VP(v,n))$ of adnominal verb phrase $VP(v,n)$ is calculated by

the following expression (1):

$$W(VP(v,n)) = \frac{E(l(n)) \times M(VP(v,n))}{idf(n) \times J(n)} \times (1 + CR(VP(v,n))), \quad (1)$$

where

$VP(v,n)$: an adnominal verb phrase having verb v modifying noun n,

$l(n)$: a noun contained at the end of compound noun n modified by adnominal verb phrases,

$E(l(n))$: entropy based on the probability that verbs modify noun $l(n)$, which will be defined by expression (3).

$idf(n)$: weight value calculated by the *idf* (Salton, 1988) method to increase the weight value of noun n dispersed in the document set. This is defined by the following expression (2).

$$idf(n) = \log \frac{|D|}{df(n)}, \quad (2)$$

where

D: the set of general documents used to acquire knowledge about the deletion possibility of adnominal verb phrases,

$df(n)$: frequency of documents containing noun n in documents set D,

$J(n)$: the number of nouns contained in compound noun n,

$M(VP(v,n))$: the number of verb phrases modifying verb v,

$CR(VP(v,n))$: a factor to reflect context of the document. This will be explained in subsection 2.3.

Step 2:

Delete adnominal verb phrase $VP(v,n)$ having weight value $W(VP(v,n))$ smaller than a threshold value predetermined by trial and error.

□

$W(VP(v,n))$ shows degree of importance of adnominal verb phrase $VP(v,n)$ and if $W(VP(v,n))$ is large, $VP(v,n)$ is not able to be deleted.

For example, $l(n)$ is "*Tougou*(統合: unification)", when noun n is "*Shijo Tougou*(市場統合: market unification)". (The "*Shijo Tougou*(市場統合: market unification)" is a compound noun composed of a noun:"*Shijo*(市場: market)" and a noun:"*Tougou*(統合: unification)")

Entropy $E(l(n))$ reflects "frequency of modification of noun $l(n)$ by adnominal verb phrases" and "variety of adnominal verb phrases modifying noun $l(n)$". By our method, "frequency of modification of noun $l(n)$ by adnominal verb phrases" and "variety of adnominal verb phrases modifying noun $l(n)$" is expressed by the entropy based on a probability that verbs modify noun $l(n)$. Entropy $E(l(n))$ has a large value if noun $l(n)$ is modified by various kinds of verbs. The verbs modified by noun $l(n)$ having a large entropy value can not be predicted easily. Similarly, it is hard to associate noun $l(n)$ having a large entropy value with an adnominal verb phrase modifying noun $l(n)$. Information loss may become serious if such adnominal verb phrases modifying noun $l(n)$ having a large entropy value are deleted. Hence, the adnominal verb phrases modifying noun $l(n)$ having a large entropy value are not deleted, because weight value $W(VP(v,n))$ increases as $E(l(n))$ increases. But, it is easy to associate noun $l(n)$ having a small entropy value with an adnominal verb phrase modifying noun $l(n)$. Consequently, the adnominal verb phrases modifying noun $l(n)$ having a small entropy value are deleted, because weight value $W(VP(v,n))$ decreases as $E(l(n))$ decreases. We will illustrate the entropy based on a probability that verbs modify a noun in subsection 2.2.

Noun n with a large $idf(n)$ value is important. Deletion of the adnominal verb phrases modifying noun n with a large $idf(n)$ value does not cause serious information loss, because meaning of such noun n is understood without any modifier. Thus, the adnominal verb phrases modifying noun n with a large $idf(n)$ value are deleted because weight value $W(VP(v,n))$ decreases as $idf(n)$ increases.

For example, when n is "*Shijo Tougou*(市場統合: market unification)", $J(n) = 2$. If noun n is not compound noun, $J(n) = 1$. If noun n

is compound noun, deletion of adnominal verb phrases modifying noun n does not cause serious information loss, because meaning of noun n is understood only by noun n. Thus, it is easy to delete adnominal verb phrases modifying such noun n, because weight value $W(VP(v,n))$ decreases as $J(n)$ increases.

$M(VP(v,n))$ is the number of verb phrases modifying verb v, which is contained in adnominal verb phrase $VP(v,n)$. $M(VP(v,n))$ is illustrated in the following example.

Example 2:
Adnominal verb phrase $VP(v,n)$: *Koremade EC shijo wo bundan site kita* (これまでＥＣ市場を分断してきた: EC market has so far been divided),
Verb v: *bundansuru* (分断する: divide),
Verb phrase 1: *koremade* (これまで: so far),
Verb phrase 2: *EC shijo* (ＥＣ市場を: EC market).

□

Value $M(VP(v,n))$ for this example is 2. If the adnominal verb phrase $VP(v,n)$ has a large $M(VP(v,n))$ value, the $VP(v,n)$ has much information. Information loss may become serious if such adnominal verb phrase $VP(v,n)$ having a large $M(VP(v,n))$ value is deleted. Consequently, the adnominal verb phrase $VP(v,n)$ with a large $M(VP(v,n))$ value is not deleted because weight value $W(VP(v,n))$ increases as $M(VP(v,n))$ increases.

2.2 The entropy based on a probability that verbs modify a noun

Our method uses "frequency of modification of a noun by adnominal verb phrases", "variety of adnominal verb phrases modifying the noun" to decide if an adnominal verb phrase is deleted. The entropy $E(l(n))$ based on a probability of verbs to modify noun $l(n)$ reflects "frequency of modification of noun $l(n)$ by adnominal verb phrases", "variety of adnominal verb phrases modifying noun $l(n)$". Entropy $E(l(n))$ is calculated by the following expression (3):

$$E(l(n)) = -\sum_{v \in V(l(n))} P(v,l(n)) \times \log(P(v,l(n))), \quad (3)$$

where

$V(l(n))$: the set of verbs contained in the adnominal verb phrases modifying noun $l(n)$ in a set of documents used to acquire knowledge,

$P(v,l(n))$: the probability that verb v modifies noun $l(n)$, in addition, $v \in V(l(n))$. The probability is calculated by the following expression (4):

$$P(v,l(n)) = \frac{freq(v,l(n))}{\sum_{v \in V(l(n))} freq(v,l(n))}, \quad (4)$$

where $freq(v,l(n))$ is the frequency of verb $v \in V(l(n))$ modifying noun $l(n)$ in the set of documents used to acquire knowledge.

2.3 A factor to reflect context of the document

$CR(VP(v,n))$ is a factor to reflect context of the document. $CR(VP(v,n))$ is calculated by the following expression (5):

$$CR(VP(v,n)) = \sum_{p \in P(VP(v,n))} B(p,VP(v,n)),$$
$$B(p,VP(v,n)) = \frac{1+after(p,VP(v,n))}{2(1+before(p,VP(v,n)))} \times \log\frac{|D|}{df(p)}, \quad (5)$$

where

$P(VP(v,n))$: the set of nouns contained in adnominal verb phrase $VP(v,n)$,

$before(p,VP(v,n))$: frequency where noun p is contained in adnominal verb phrase $VP(v,n)$ which appears before $VP(v,n)$ in the document.

$after(p,VP(v,n))$: frequency where noun p is contained in adnominal verb phrase $VP(v,n)$ which appears after $VP(v,n)$ in the document.

$CR(VP(v,n))$ has small value if noun p is contained in adnominal verb phrase $VP(v,n)$ which appears before $VP(v,n)$ in a document. Thus, adnominal verb phrase $VP(v,n)$ containing noun p appearing before $VP(v,n)$ in a document tends to be deleted by our method, because weight value $W(VP(v,n))$ decreases as $CR(VP(v,n))$ decreases.

If important nouns are contained in adnominal verb phrase $VP(v,n)$, $VP(v,n)$ does not tend to be deleted, because, adnominal verb

phrases containing the important nouns have important contents. $CR(VP(v,n))$ has a large value if noun p contained in adnominal verb phrase $VP(v,n)$ has important contents. Note that if noun p has important contents, $df(p)$ has small value. Thus, adnominal verb phrase $VP(v,n)$ containing noun p having important contents is not deleted by our method, because weight value $W(VP(v,n))$ increases as $CR(VP(v,n))$ increases.

2.4 Constraints on deletion for improving precision

We impose the following constraints on deletion for improving precision.

Constraint 1: Adnominal verb phrases where $CR(VP(v,n)) = 1$ are not able to be deleted.

Constraint 2: if adnominal verb phrase $VP(v1,n1)$ contains a first noun modified by another adnominal verb phrase $VP(v2,n2)$ in $VP(v1,n1)$, $VP(v1,n1)$ is not able to be deleted.

Constraint 3: Adnominal verb phrases modifying a noun listed in Table 1, are not able to be deleted.

The case of $CR(VP(v,n)) = 1$ occurs when an adnominal verb phrase has no noun. If an adnominal verb phrase having no noun is deleted, the cohesion among sentences may be destroyed. Thus, we impose Constraint 1.

The reason why we impose Constraint 2 is to prevent excessive deletion. That is, without this constraint, if adnominal verb phrase $VP(v1,n1)$ is deleted, the $VP(v2,n2)$ must be also deleted.

Nouns having no substantial meaning, e.g., the nouns listed in Table 1, is used for Constraint 3. Note that some Japanese nouns in the table do not correspond to English nouns when translated.

Table 1: The nouns list for Constraint 3

koto(こと: thing)	mono(もの: thing)
wake(わけ: reason)	ue(上: upper)
naka(中: among)	hoka(他: other)
hoka(ほか: other)	mae(前: before)
ato(後: after)	aida(間: between)
sai(際: occasion)	ue(うえ: upper)
tame(ため: reason)	kurai(くらい: about)
tokoro(ところ: place)	you(よう: like)
kagiri(かぎり: limit)	hituyou(必要: need)
ugoki(動き: movement)	ippou(一方: side)
sita(下: under)	muki(向き: direction)
naka(なか: among)	

3 Implementation

We implemented our method. We use 66686 documents from Nikkei newspaper articles from January 1, to June 31, 1993, as a document set. We employ JUMAN[1] version 3.5 as a morphological analyzer, and KNP[2] version 2.0b6 as a parser. Our method acquires knowledge about the deletion possibility of adnominal verb phrases, and deletes adnominal verb phrases by using the acquired knowledge. The adnominal verb phrases deleted by our method are exemplified as follows. Note that the adnominal verb phrases deleted by our method are underlined.

Example 3:
Kyanon ha ryousan ka no nekku ni natteita taiyoudenchi no maku seisei sokudo wo hiyakuteki ni takameru gijutsu wo kaihatsu shi... (キヤノンは量産化のネックになっていた太陽電池の膜生成速度を飛躍的に高める技術をこのほど開発し、...: Canon has recently developed a technology which drastically increases the film generation speed of a solar cell which had been the bottleneck of mass-production.)

Example 4: *Kankyou mondai wo haikei ni shita tennen gas no zyuyou kakudai de, kaigai de no LNG kaihatu sannyuu ni hazumi ga tuiteiru* (環境問題を背景にした天然ガスの需要拡大で、海外でのＬＮＧ開発参入に弾みがついている。: Entry to LNG development is promoted overseas, because demand on the natural gas expands by the reason that the environmental problem has become more seriously concerned.)

4 Experiments for evaluation

We evaluate our method by experiments. We choose 20 among the 66686 documents for deleting adnominal verb phrases by our method. There are 174 adnominal verb phrases in the 20 documents.

[1] http://www-lab25.kuee.kyoto-u.ac.jp/nl-resource/juman.html

[2] http://www-lab25.kuee.kyoto-u.ac.jp/nl-resource/knp.html

Table 2: The results of the experiment by our method

Threshold	Recall(%)	Precision(%)	F measure	Deleted
3.6	68.5	79.4	73.5	63
3.8	72.6	80.3	76.3	66
4	72.6	79.1	75.7	67
4.1	74.0	79.4	76.6	68
4.2	74.0	79.4	76.6	68
4.4	74.0	78.3	76.1	69
4.6	74.0	78.3	76.1	69
Average	72.7	79.3	75.8	67.0

We evaluate our method by precision and recall. For this purpose, we manually make a correct data set which shows adnominal verb phrases appropriate to be deleted among the 174 adnominal verb phrases. Note that the precision and the recall are defined as follows.

$$\begin{aligned} Recall\ R &= freq(A)/freq(C), \\ Precision\ P &= freq(A)/freq(M), \\ F\ measure\ F &= \frac{2RP}{R+P}, \end{aligned}$$

where,

freq(A): the frequency of the same adnominal verb phrases shown by the correct data set with adnominal verb phrases deleted by our method,

freq(C): the frequency of adnominal verb phrases shown by the correct data set,

freq(M): the frequency of adnominal verb phrases deleted by our method.

Precision and recall are calculated for threshold values changed from 3.6 to 4.6. Table 2 shows a part of the results. The reason why the threshold value is changed from 3.6 to 4.6 is that the F measure is best when the threshold value is 4.1.

Next, we compare results of the following three methods for evaluation.

Method A: our method.

Method B: the method where adnominal verb phrases which do not match the Constraints 1, 2, 3 shown in subsection 2.4 are all deleted.

Table 3: The result of each other

Method	A	B	C
Deleted	67.0	134	81
Recall(%)	72.7	90.4	72.6
Precision(%)	79.3	49.3	65.4

Method C: the method obtained from our method by removing $CR(VP(v,n))$ from expression (1).

We employ the average of precision and that of recall for evaluating Method A.

Method B only uses information on the syntactic structure of a sentence. We evaluate Method A by comparing with Method B.

Moreover, we evaluate effect of $CR(VP(v,n))$ by comparing with Method C. The threshold value is adjusted so that the recall attained by the Method A may become the same value. The result is shown in Table 3.

Next, we calculate the average of reduction rate which results from the deletion of adnominal verb phrases from the sentences. Note that the reduction rate R_{rate} is defined as follows:

$$R_{rate} = W(reduction)/W(source),$$

where,

W(source): Number of words constructing a sentence containing adnominal verb phrases which can be deleted by our method;

W(reduction): Number of words constructing the sentence where the adnominal verb phrases have been deleted by our method.

We calculate R_{rate} of all sentences containing adnominal verb phrases which can be deleted by our method in the 20 documents. And we calculate the average of R_{rate}. Thus, R_{rate} is 79.3%

5 Discussion

We consider that our method is able to delete adnominal verb phrases appropriately, because the precision and the recall attain 79.3% and 72.7%, respectively. And we conclude that Method A is superior to Method B, because the precision of Method A is higher than that of Method B. Experimental results show that our method is useful for deleting adnominal verb phrases.

Moreover, we conclude that $CR(VP(v,n))$ is useful for this task, because the precision and the recall of Method A are higher than those of Method C. The correct adnominal verb phrase deleted by our method is as follows.

Example 5: *Kantan ni rokuga yoyaku ga dekiru G koudo naizou gata no VTR wo tounyuu site iru ...* (簡単に録画予約ができるＧコード内蔵型のＶＴＲを投入している... : G code built-in type VTR which is able to reserve videotape-recording is thrown in...)

The adnominal verb phrase: "*Kantan ni rokuga yoyaku ga dekiru*(簡単に録画予約ができる: be able to reserve videotape-recording)" modifying the noun "*G koudo naizou*(Ｇコード内蔵: G code built-in)" is deleted by our method. We consider that this adnominal verb phrase is able to be deleted, because we can easily supplement the "*Kantan ni rokuga yoyaku ga dekiru*(簡単に録画予約ができる: be able to reserve videotape-recording)" from general knowledge. The reason why our method deletes it is that the weight value of "*Kantan ni rokuga yoyaku ga dekiru*(簡単に録画予約ができる: be able to reserve videotape-recording)" is small, because the entropy value of noun "*naizou*(内蔵: built-in)" is small. (The entropy value of noun "*naizou*(内蔵: built-in)" is 0.693. In contrast, the entropy value of noun "*an*(案: plan)" is 4.13.) The noun "*naizou*(内蔵: built-in)" is seldom modified by adnominal verb phrases. Consequently, the entropy value of noun "*naizou*(内蔵: built-in)" is small. It is easy to associate a noun having a small entropy value with an adnominal verb phrase modifying the noun. Thus, the adnominal verb phrases modifying the noun having a small entropy value are deleted.

The incorrect deletion of an adnominal verb phrase by our method is illustrated in the following example.

Example 6: *Doitsu de ha koremade nandoka, kousoku-douro yuuryouka an ga detaga, kokunai, kokugai no hantai de tsubureteki ta ikisatsu ga aru.* (ドイツではこれまで何度か、高速道路有料化案が出たが、国内、国外の反対でつぶれてきたいきさつがある。: In Germany, although charging highway traffic is proposed several times, there are circumstances by which such proposals have been crushed by objection from the domestic and foreign people.)

We consider that the adnominal verb phrase "*kokunai, kokugai no hantai de tsubureteki ta*(国内、国外の反対でつぶれてきた: have been crushed by objection from the domestic and foreign people)" is not able to be deleted, because the noun "*ikisatsu*(いきさつ: circumstances)" has no substantial meaning. However, the noun "*ikisatsu*(いきさつ: circumstances)" is recognized as an important noun by our method, because the noun "*ikisatsu*(いきさつ: circumstances)" has a large *idf* value. Hence, the weight value of this adnominal verb phrase is small, and it is deleted by our method. If a noun "*keii*(経緯: circumstances)" which has the same meaning with "*ikisatsu*(いきさつ: circumstances)" is modified by the adnominal verb phrase, the adnominal verb phrase may not be deleted. Because the noun "*keii*(経緯: circumstances)" has a small *idf* value and has a large entropy value. (The entropy value of noun "*ikisatsu*(いきさつ: circumstances)" is 1.61. On the other hand, the entropy value of noun "*keii*(経緯: circumstances)" is 4.17.) We consider that this problem is solvable by adding constraint that if a noun "A" with a similar meaning of a noun "B"to which our method is applied exists and noun "A" has a small *idf* value and a large entropy value. The adnominal verb phrase modified by noun "B" is not deleted.

6 Conclusion

We proposed an unsupervised learning method which acquires knowledge about the deletion possibility of adnominal verb phrases from news

paper corpus provided in a machine-readable form in order to summarize by deleting unimportant segment from a sentence. Actually, our method deletes adnominal verb phrases modifying nouns not frequently modified by verbs, not modified by many kinds of verbs. We evaluate our method, and we conclude that our method is able to delete adnominal verb phrases appropriately, because the precision and the recall attain 79.3% and 72.7%, respectively. Experimental results show that our method is useful to delete adnominal verb phrases.

Acknowledgment

We express our gratitude to Nikkei Shinbun, Inc. who permits us to use the Nikkei news paper articles in a machine readable form.

References

H. Jing and K. McKeown. 2000. Cut and paste based text summarization. In *Proc. of the 1st Meeting of the North American Chapter of the Association for Computational Linguistics*, pages 178–185.

H. Jing. 2000. Sentence reduction for automatic text summarization. In *Proc. of the 6th Conference on Applied Natural Language Processing*, pages 310–315.

N. Katoh and N. Uratani. 1999. A new approach to aquiring linguistic knowledge for locally summarizing japanese news sentences. *Natural Language Processing*, 6(7):73–92 (in Japanese).

K. Knight and D. Marcu. 2000. Statistics-based summarization –step one: Sentence compression. In *Proc. of AAAI2000*, pages 703–710.

I. Mani and M. T.Maybury. 1999. *Advances in Automatic Text Summarization.* the MIT Press.

K. Ohtake, D. Okamoto, M. Kodama, and S. Masuyama. 2001. Yet another summarization system with two modules using empirical knowledge. In *Proceedings of the Second NTCIR Workshop Meeting on Evaluation of Chinese & Japanese Text Retrieval and Text Summarization*, pages 225–234.

G. Salton. 1988. *Automatic Text Processing.* Addison Wesley.

K. Takeuchi and Y. Matsumoto. 2001. Acquistion of sentence reduction rules for improving quality of text summaries. In *Proceedings of the Sixth Natural Language Processing Pacific Rim Symposium*, pages 447–452.

K. Yamamoto, S. Masuyama, and S. Naito. 1995. Green: An experimental system generating summary of japanese editorials by combining multiple discourse characteristics. *Natural Language Processing*, 2(1):39–55 (in Japanese).