Producing More Readable Extracts by Revising Them

Hidetsugu Nanba (†) and **Manabu Okumura** (†, ‡)
†School of Information Science
Japan Advanced Institute of Science and Technology
‡Precision and Intelligence Laboratory
Tokyo Institute of Technology
nanba@jaist.ac.jp, oku@pi.titech.ac.jp

Abstract

In this paper, we first experimentally investigated the factors that make extracts hard to read. We did this by having human subjects try to revise extracts to produce more readable ones. We then classified the factors into five, most of which are related to cohesion, after which we devised revision rules for each factor, and partially implemented a system that revises extracts.

1 Introduction

The increasing number of on-line texts available has resulted in automatic text summarization becoming a major research topic in the NLP community. The main approach is to extract important sentences from the texts, and the main task in this approach is that of evaluating the importance of sentences [MIT, 1999]. This producing of extracts - that is, sets of extracted important sentences - is thought to be easy, and has therefore long been the main way that texts are summarized. As Paice pointed out, however, computer-produced extracts tend to suffer from a 'lack of cohesion' [Paice, 1990]. For example, the antecedents corresponding to anaphors in an extract are not always included in the extract. This often makes the extracts hard to read.

In the work described in this paper, we therefore developed a method for making extracts easier to read by revising them. We first experimentally investigated the factors that make extracts hard to read. We did this by having human subjects try to revise extracts to produce more readable ones. We then classified the factors into five, most of which are related to cohesion [Halliday et al., 1976], after which we devised revision rules for each factor, and partially implemented a system that revises extracts. We then evaluated our system by comparing its revisions with those produced by human subjects and also by comparing the readability judgments of human subjects between the revised and original extracts.

In the following sections we briefly review related works, describe our investigation of what make extracts hard to read, and explain our system for revising extracts to make them more readable. Finally, we describe our evaluation of the system and discuss the results of that evaluation.

2 Related Works

Many investigators have tried to measure the readability of texts [Klare, 1963]. Most of them have evaluated well-formed texts produced by people, and used two measures: percentage of familiar words in the texts (word level) and the average length of the sentences in the texts (syntactic level). These measures, however, do not necessarily reflect the actual readability of computer-produced extracts. We therefore have to take into account other factors that might reduce the readability of extracts.

One of them could be a lack of cohesion. Halliday and Hasan [Halliday et al., 1976] described five kinds of cohesion: reference, substitution, ellipsis, conjunction, and lexical cohesion.

Minel [Minel et al., 1997] tried to measure the readability of extracts in two ways: by counting the number of anaphors in an extract that do not have antecedents in the extract, and by counting the number of sentences which are not included in an extract but closely connected to sentences in the extract.

We therefore regard kinds of cohesion as important in trying to classify the factors that make extracts less readable in the next section.

One of the notable previous works dealing with ways to produce more cohesive extracts is that of Paice [Paice, 1990]. Mathis presented a framework in which a pair of short sentences are combined into one to yield a more readable extract [Mathis et al., 1973]. We think, however, that none of the previous studies have adequately investigated the factors making extracts hard to read.

Some investigators have compared human-produced abstracts with the original texts and investigated how people revise texts to produce abstracts

[Kawahara, 1989, Jing, 1999]. Revision is thought to be done for (at least) the following three purposes:

(1) to shorten texts,

(2) to change the style of texts,

(3) to make texts more readable.

Jing [Jing, 1999] is trying to implement a human summarization model that includes two revision operations: reduction (1) and combination (3). Mani [Mani et al., 1999] proposed a revision system that uses three operations: elimination (1), aggregation (1), and smoothing (1, 3). Mani showed that his system can make extracts more informative without degrading their readability. The present work, however, is concerned not with improving readability but with improving the informativeness.

3 Less Readability of Extracts

To investigate the revision of extracts experimentally, we had 12 graduate students produce extracts of 25 newspaper articles from the NIHON KEIZAI SHINBUN, the average length of which was 30 sentences. We then asked them to revise the extracts (six subjects per extract).

We obtained extracts containing 343 revisions, made for any of the three purposes listed in the last section. We selected the revisions for readability, and classified them into 5 categories, by taking into account the categories of cohesion by Halliday and Hasan[Halliday et al., 1976]. Table 1 shows the sum of the investigation.

Next, we illustrate each category of revisions. In the examples, darkened sentences are those that are not included in extracts, but are shown for explanation. The serial number in the original text is also shown at the beginning of sentences [1].

A) Lack of conjunctive expressions/presence of extraneous conjunctive expressions

The relation between sentences 15 and 16 is adversative, because there is a conjunctive 'しかし (However)' at the beginning of sentence 16. But because sentence 15 is not in the extract, 'しかし (However)' is considered unnecessary and should be deleted. Conversely, lack of conjunctive expressions might cause the relation between sentences to be difficult to understand. In such a case, a suitable conjunctive expression should be added. For these tasks, discourse structure analyzer is required.

---

[1] We use the following three tags to show revisions.
< add > E_1 < /add >: add a new expression E_1.
< del > E_2 < /del >: delete an expression E_2.
< rep E_4 > E_3 < /rep >: replace an expression E_3 with E_4.

3. 専任社員を置くことで女性の職域を積極的に拡大していく方針だ.
(The company plans to give women more opportunity to work by employing full-time workers.)

15. 会社の期待は大きいが，このような女性参加のプロジェクトの前例がなく，現在は悪戦苦闘中.
(Since there have been no similar cases before, the project that women join is now in a hard situation, though the company puts hopes on it.)

16. <del>しかし </del>目標を二〇〇一年に置いて女性社員にとって満足できる仕事環境，会社にとって有効な人材の活用と両者に利益のある改革を進めている.
(~~However,~~ it is making efforts of reformation which will be profitable both for the company and the female workers.)

B) Syntactic complexity

2. (修正前;before revision)
樫尾和雄社長自ら「将来の経営の核」に挙げる通信事業の第一弾であり，今後，携帯電話などに手を広げていくための布石でもある.
(It is the first project in telecommunication business, which President Kashio wants to be one of the central businesses in the future, and it is also the preparation for expanding the business to cellular phone.)

↓

(修正後;after revision)
樫尾和雄社長自ら「将来の経営の核」に挙げる通信事業の第一弾である.
(It is the first project in telecommunication businesses, which President Kashio wants to be one of the central business in the future.)
今後，携帯電話などに手を広げていくための布石でもある.
(It is also the preparation for expanding the business to cellular phone.)

Longer sentences tend to have a syntactically complex structure [Klare, 1963], and a long compound sentence should generally be divided into two simpler sentences. It has also been claimed, however, that short coordinate sentences should be combined [Mathis et al., 1973].

C) Redundant repetition

2. 越後製菓が四月に始めた新商品シリーズ「越後ベスト一〇〇」が主婦に人気だ.
(The new product 'ECHIGO BEST 100' which ECHIGO SEIKA released this April is popular among housewives.)

4. <rep 同社 > 越後製菓 </rep> は米菓分野で八七年からNTT のキャプテンなどを活用.
(<rep The company> ECHIGO SEIKA </rep> has been making use of NTT Captain system since 1987.)

If subjects of adjacent sentences in an extract are the same, as in the above example, readers might think they are redundant. In such a case, repeated expressions should be omitted or replaced by pronouns. In this example, the anaphoric expression '同社 (the company)' is used instead of the original expression.

Table 1: Factors of less readability and their revision methods

	factors	revision methods	required techniques
A	lack of conjunctive expressions/ presence of extraneous conjunctive expressions	add/delete conjunctive expressions	discourse structure analysis
B	syntactic complexity	combine two sentences; divide a sentence into two	
C	redundant repetition	pronominalize; omit expressions; add demonstratives	
D	lack of information	supplement omitted expressions; replace anaphors by antecedents; delete anaphors	anaphora and ellipsis resolution
		add supplementary information	information extraction
E	lack of adverbial particles; presence of extraneous adverbial particles	add/delete adverbial particles	

D) Lack of information

> 2. 自動車のクライスラーとパソコンのコンパック・コンピューターだ.
> (These are the car maker CHRYSLER and the computer maker COMPAC.)
>
> 8. 企業の人員削減が消費意欲を沈滞させ，それがまた企業業績の足を引っ張るというデフレ型の悪循環に陥りつつある.
> (We are now in a vicious circle where the layoffs by companies discourage consumptions, which in turn results in lower sales.)
>
> 9. <del> その中で，</del> クライスラーが大健闘しているのは，新製品投入の時期と価格・商品戦略がまさしく低成長時代に適応しているからだ.
> (~~In such a situation,~~ CHRYSLER has done well, because its management strategy exactly fits the age of low growth.)

In this example, the referent of 'その中で (in such a situation)' in sentence 9 is sentence 8, which is not in the extract. In such a case, there are two ways to revise: to replace the anaphoric expression with its antecedent, or to delete the expression. The revision in the example is the latter one. For the task, a method for anaphora and ellipsis resolution is required.

> 1. ソフトバンクの孫正義社長はこのところ時差ボケが直らない.
> (Masayoshi Son, CEO of Softbank, is now suffering from jet lag.)
>
> 3. <add> ソフトバンクの </add> 孫社長が「社運をかけたプロジェクト」と意気込む新流通システムは CD-ROM を使ったソフトの販売.
> (CEO Son <add> of Softbank </add> is eager to sell softwares using CD-ROM, and he think it is a big project for his company.)

In this second example, since 'CEO Son' appears without the name of the company in the extract, without any background knowledge, we may not understand what company Mr. Son is the CEO of. Therefore, the name of the company 'Softbank' should be added as the supplementary information. The task requires a method for information extraction or at least named entity extraction.

E) lack of adverbial particles/presence of extraneous adverbial particles

> 26. 今回のド・ムオイ書記長の訪日は両国が相互理解を深めるうえで，よい機会である.
> (It is a good opportunity to promote the mutual understanding between Japan and Vietnam that Mr. Do MUOI, a chief secretary of Vietnam, visits Japan.)
>
> 29. ベトナムはアジアの安全保障の観点からも今後，重要性を増していく.
> (From a viewpoint of security, Vietnam will be a key country in Asia.)
>
> 30. 経済協力の面で <del> も </del> 長い目で発展を支援していく姿勢が必要だろう.
> (Japanese government should consider long-term economical support<del>, too </del>.)

In the above example, there is an adverbial particle 'も (, too)' and we can find that sentences 29 and 30 are paratactical. But, because sentence 29 is not in the extract, the particle 'も (, too)' is unnecessary and should be deleted.

4 Revision System

Our system uses the Japanese public-domain analyzers JUMAN [Kurohashi et al., 1998] and KNP [Kurohashi, 1998] morphologically and syntactically analyze an original newspaper article and its extract. It then applies revisions rules to the extract repeatedly, with reference to the original text, until no rules can revise the extract further.

4.1 Revision Rules

Because the techniques needed for dealing with all the categories of revisions dealt with in the previous

section were not available, we devised and implemented revision rules only for factors (A), (C), and (D) in Table 1 by using JPerl.

a) Deletion of conjunctive expressions

We prepared a list of 52 conjunctive expressions, and made it a rule to delete each of them whenever the extract does not include the sentence that expression is related. To identify the sentence related to the sentence by the conjunction [Mann et al., 1986], the system performs partial discourse structure analysis taking into account all sentences within three sentences of the one containing the conjunctive expression.

The implementation of our partial discourse structure analyzer was based on Fukumoto's discourse structure analyzer [Fukumoto, 1990]. It infers the relationship between two sentences by referring to the conjunctive expressions, topical words, and demonstrative words.

c) Omission of redundant expressions

If subjects (or topical expressions marked with topical postposition 'wa') of adjacent sentences in an extract were the same, the repeated expressions were considered redundant and were deleted.

d-1) Deletion of anaphors

To treat anaphora and ellipsis successfully, we would need a mechanism for anaphora and ellipsis resolution (finding the antecedents and omitted expressions). Because we have no such mechanism, we implement a rule with *ad hoc* heuristics: If an anaphor appears at the beginning of a sentence in an extract, its antecedent must be in the preceding sentence. Therefore, if that sentence was not in the extract, the anaphor was deleted.

d-2) Supplement of omitted subjects

If a subject in a sentence in an extract is omitted, the revision rule supplements the subject from the nearest preceding sentence whose subject is not omitted in the original text. This rule is implemented by using heuristics similar to the above revision rule.

5 Evaluation of Revision System

We evaluated our revision system by comparing its revisions with those by human subjects (evaluation 1), and comparing readability judgments between the revised and original extracts (evaluation 2).

5.1 Evaluation 1: comparing system revisions and human revisions

Because revision is a subjective task, it was not easy to prepare an answer set of revisions to which our system's revisions could be compared. The revisions that more subjects make, however, can be considered more reliable and more likely to be necessary. When comparing the revisions made by our system with those made by human subjects, we therefore took into account the degree of agreement among subjects.

For this evaluation, we used 31 newspaper articles (NIHON KEIZAI SHINBUN) and their extracts. They were different from the articles used for making rules. Fifteen of extracts are taken from Nomoto's work [Nomoto et al., 1997], and the rest were made by our group. The average numbers of sentences in the original articles and the extracts were 25.2 and 5.1.

Each extract was revised by five subjects who had been instructed to revise the extracts to make them more readable and had been shown the 5 examples in section 3. As a result, we obtained 167 revisions in total. The results are listed in Table 2.

Table 2: The number of revisions

	revision methods	total
A	add(61)/delete(11) conjunctive expressions	72
B	combine two sentences(2) divide a sentence into two(6)	8
C	pronominalize(5); omit expressions(3) add demonstratives(8)	16
D	supplement omitted expressions(11) replace anaphors by antecedents(10) delete anaphors(15)	36
	add supplementary information(26)	26
E	delete adverbial particles(4) add adverbial particles(5)	9
		167

We compared our system's revisions with the answer set comprising revisions that more than two subjects made. And we used recall (R) and precision (P) as measures of the system's performances.

$$R = \frac{\left(\begin{array}{c} Number\ of\ system's\ revisions \\ matched\ to\ the\ answer \end{array} \right)}{Number\ of\ revisions\ in\ the\ answer}$$

$$P = \frac{\left(\begin{array}{c} Number\ of\ system's\ revisions \\ matched\ to\ the\ answer \end{array} \right)}{Number\ of\ system's\ revisions}$$

Evaluation results are listed in Table 3. As in Table 3, the coverage of our revision rules is rather small (about 1/4) in the whole set of revisions in Table 2. It is true that the experiment is rather small and can be considered as less reliable. Though it is less reliable, some of the implemented rules can cover most of the necessary revisions by human subjects. However, precision should be improved.

Table 3: Comparison between the revisions by human and our system

revision rules	R	P
a(total:11)	2/2	2/5
c(total:3)	0/0	0/0
d-1(total:15)	4/5	4/7
d-2(total:11)	2/4	2/10

5.2 Evaluation 2: comparing human readability judgments of original and revised extracts

In the second evaluation, using the same 31 texts as in evaluation 1, we asked five human subjects to rank the following four kinds extracts in the order of readability: the original extract (without revision)(NON-REV), human-revised ones (REV-1 and REV-2), and the one revised by our system (REV-AUTO). REV-1 and REV-2 were respectively extracts revised in the cases where more than one and more than two subjects agreed to revise.

We considered a judgment by the majority (more than two subjects) to be reliable. The results are listed in Table 4. The column 'split' in Table 4 indicates the number of cases where no majority could agree. The results show that both REV-1 and REV-2 extracts were more readable than NON-REV extracts and that REV-2 extracts might be better than REV-1 extracts, since the number of 'worse' evaluations was smaller for REV-2 extracts.

Table 4: Comparison of readability among original extracts and revised ones

	better	same	worse	split
REV-2 vs. NON	15	12	2	2
REV-1 vs. NON	22	1	7	1
AUTO vs. NON	2	13	12	0

In comparing REV-AUTO with NON-REV, we use 27 texts where the readability does not degrade in REV-2, since the readability cannot improve with revisions by our system in those texts where the readability degrades even with human revisions. Even with those texts, however, in almost half the cases, the readability of the revised extract was worse than that of the original extract. The main reason is that the revision system supplemented incorrect subjects.

6 Discussion

Although the results of the evaluation are encouraging, they also show that our system needs to be improved. We have to implement more revision rules to enlarge the coverage of our system. One of the most frequent revisions is to add conjunctions(37%). We also need to reform our revision rules into more thorough implementation. To improve our system,

we think it is necessary to develop a robust discourse structure analyzer, a robust mechanism for anaphora and ellipsis resolution, and a robust system of extracting named entities. They are under development now.

7 Conclusion

In this paper we described our investigation of the factors that make extracts less readable than they should be. We had human subjects revise extracts to made them more readable, and we classified the factors into five categories. We then devised revision rules for three of these factors and implemented a system that uses them to revise extracts. We found experimentally that our revision system can improve the readability of extracts.

References

[Fukumoto, 1990] Fukumoto, J. (1990) Context Structure Analysis Based on the Writer's Insistence. *IPSJ SIG Notes, NL-78-15, pp.113–120.* in Japanese.

[Halliday et al., 1976] Halliday,M.A.K., Hasan,R. (1976) Cohesion in English. Longman.

[Jing, 1999] Jing,H. (1999) Summary Generation through Intelligent Cutting and Pasting of the Input Document. *Ph.D. Thesis Proposal, Columbia Univ.*

[Kawahara, 1989] Kawahara,H. (1989) Chapter 9, in Bunshoukouzou to youyakubun no shosou. *Kuroshio-shuppan,* pp.141–167. in Japanese.

[Klare, 1963] Klare,G.R. (1963) The Measurement of Readability. *Iowa State University Press.*

[Kurohashi et al., 1998] Kurohashi,S., Nagao,M. (1998) Japanese Morphological Analysis System JUMAN version 3.5.

[Kurohashi, 1998] Kurohashi, S. (1998) Japanese parser KNP version 2.0 b6.

[Mathis et al., 1973] Mathis,B., Rush,J., Young,C. (1973) Improvement of Automatic Abstracts by the Use of Structural Analysis. *JASIS,*24(2),pp.101–109.

[Mani et al., 1999] Mani,I., Gates,B., Bloedorn,E. (1999) Improving Summaries by Revising Them. *the 37th Annual Meeting of the ACL,* pp.558–565.

[MIT, 1999] Mani,I., Maybury,M.T. (1999) Advances in Automatic Text Summarization. MIT Press.

[Mann et al., 1986] Mann,W.C., Thompson,S.A. (1986) Rhetorical Structure Theory: Description and Construction of Text Structure. *Proc. of the third International Workshop on Text Generation.*

[Minel et al., 1997] Minel,J., Nugier,S., Gerald,P. (1997) How to Appreciate the Quality of Automatic Text Summarization? Examples of FAN and MLUCE Protocols and their Results on SERAPHIN. *Intelligent Scalable Text Summarization, Proc. of a Workshop, ACL,* pp.25–30.

[Nomoto et al., 1997] Nomoto,T., Matsumoto,Y. (1997) The Readability of Human Coding and Effects on Automatic Abstracting. *IPSJ SIG Notes,* NL-120-11, pp.71–76. in Japanese.

[Paice, 1990] Paice,C.D. (1990) Constructing Literature Abstracts by Computer: Techniques and Prospects. *Info. Proc. & Manage.,* **26** (1), pp. 171–186.