Systems and Computers in Japan, Vol. 34, No. 12, 2003
Translated from Denshi Joho Tsushin Gakkai Ronbunshi, Vol. J85-D-II, No. 7, July 2002, pp. 1201–1210

Compilation of Dictionaries for Semantic Attribute Analysis of Television News Captions

Ichiro Ide,[1] Reiko Hamada,[2] Shuichi Sakai,[3] and Hidehiko Tanaka[3]

[1]National Institute of Informatics, Tokyo, 101-8430 Japan

[2]Graduate School of Engineering, The University of Tokyo, Tokyo, 113-8656 Japan

[3]Graduate School of Information Science and Technology, The University of Tokyo, Tokyo, 113-0033 Japan

SUMMARY

With the increase in the amount of video that is broadcast daily, there is an increasing need for storage of video in a systematic way for future reuse and retrieval. In particular, from the viewpoint of importance and usability, it is desirable to index news videos. For adequate automatic indexing based on the text information in the video, it is not sufficient to apply the simple index extraction and annotation methods which have been widely used in conventional methods. It is important to select index candidates with reference to semantic attributes. The purpose of this study is to compile dictionaries which are needed for analyzing the semantic attributes of captions (noun phrases) in TV news videos. We describe the process by which words are extracted from text corpora and a thesaurus for storage on the basis of specified conditions. The quality of the dictionaries is examined by analysis of the semantic attributes of the words appearing in actual news videos, and the results are presented. In evaluation experiments in which an existing proper noun dictionary and temporal noun dictionary were combined and used, a recall of 79 to 93% and a precision of 41 to 71% were obtained. Although the precision is low in this result, it is concluded that the compiled dictionaries are of practical use for indexing since the recall is more important in that case. © 2003 Wiley Periodicals, Inc. Syst Comp Jpn, 34(12): 32–44, 2003; Published online in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/scj.10417

Key words: indexing; caption; semantic attribute; suffix noun; dictionary.

1. Introduction

With the increasing amount of broadcast video, there is an increasing need to store videos in a systematic way for reuse and retrieval. In particular, from the viewpoint of the importance of their contents and their usability it is desirable to index news videos. At present, this procedure is performed manually in an approximate way, but automation of the process is desirable in order to achieve more precise indexing that meets the requirements for detailed retrieval.

The objective of the authors is to realize automatic indexing of the news videos by comprehensively utilizing image information and text information. In order to handle the requirements for detailed retrieval, the use of simple and approximate indexing, such as is applied in most of the existing methods, is not sufficient, and the analysis of semantic attributes must be emphasized in the selection of index candidates.

As prominent in the News-on-Demand system [17] in the Informedia Project [18] at Carnegie Mellon Univer-

sity, there are various efforts to perform automatic indexing of news videos by integrated media processing. Most of these, however, are simple indexing methods based on statistical information such as term frequencies, or are based on the occurrence timing of words or phrases. Correspondence between the index and the image content is not guaranteed, although the requirement for automatic indexing is satisfied to some extent. These methods can be used in rough indexing for each topic, but in more detailed indexing, such as indexing for each shot, the fit between the image and the index in the corresponding range may not be adequate.

In this context, the authors undertook to construct an automatic indexing system as shown in Fig. 1, considering the correspondence between the index and the image content in terms of attributes. In the figure, the index is determined with reference to correspondence between the image and the text with the same attributes. In the natural language processor used in this system, the semantic attributes must be analyzed for the text that is the index candidate. For this purpose, dictionaries were compiled in this study for use in the semantic attribute analysis of superimposed captions in news videos. This paper describes the construction process, and evaluates the performance of the dictionary by analysis of actual news captions.

More precisely, the analysis of semantic attributes in this study consists of the classification of noun phrases in terms of the following four semantic attributes: (1) person, (2) location or organization, (3) time, and (4) other. The evaluation presented in this paper considers the analysis of news captions, but the semantic attribute suffix noun dictionaries that were constructed in this study will be applied to information retrieval such as extraction of proper expressions, to other similar indexing procedures [6, 12, 15], and to general natural language understanding procedures. In the following, Section 2 describes the method of analysis of semantic attributes. Section 3 describes the compilation of the dictionary. Section 4 presents an evaluation based on application to actual news captions. Section 5 presents a summary.

2. Analysis of News Captions with Reference to Suffix Nouns

This section describes the properties of news captions and presents the method of analysis of semantic attributes using the dictionary. The video contains various kinds of text information, such as speech and closed caption texts in addition to the ordinary captions (open captions). These information sources generally contain various kinds of information, but since their content is redundant, it is necessary to extract important words and phrases.

Furthermore, although the performance of speech recognition technology is improving, it has not reached the level at which speech can be used as an accurate text information source. Closed captions have the possibility to make up for this defect of speech recognition technology. In fact, there is a widespread use of such captions in Europe and the United States. However, Japan has just begun to provide the technique to a limited extent via speech recognition limited to particular speakers (announcers) [1]. It has not been developed to a stage at which the technique is fully utilized.

On the other hand, the caption represents important information in a straightforward way, and the process of extracting the important words is not needed as it is if other information sources are used. One problem is that when the caption is superimposed on a background image, it is difficult to apply existing OCR (optical character recognition) techniques, since the resolution of the characters is degraded as a result of the small number of scan lines (525 in NTSC broadcasts). Consequently, there have been several studies focusing on the recognition of caption characters [3, 9], which have had some success.

For these reasons, this paper considers the direct use of captions as the text information source for indexing, and discusses the compilation of the dictionaries which are needed for the analysis. Although captions written down manually were used in this experiment, in the future it is planned to utilize the results of the above-mentioned studies and data acquired from data broadcasts.

2.1. Linguistic properties of captions

The caption has special properties, from the viewpoints of both grammar and content, in contrast to other general texts. From the viewpoint of grammar, almost half of the caption text consists of noun phrases, in the form of isolated substantives and noun sequences. Captions representing persons, places or organizations, and time which are to be classified in terms of the attributes are mostly noun phrases. Consequently, the noun phrase is taken as the object of analysis. From the viewpoint of content, captions representing person, place/organization, and time account for 48.3% of all captions, and there are 0.39 such noun phrases per shot on average.* Thus, a relatively large amount of information is acquired for indexing.

2.2. Semantic attribute analysis of noun phrases

Considering the linguistic properties described in the previous section, captions composed of noun phrases con-

*This is a result obtained when 2842 captions occurring in 370 minutes of news videos were analyzed manually.

cerning (1) persons, (2) locations or organizations, and (3) time were analyzed. The reason for not separating places and organizations, which are usually handled separately in this kind of analysis, is as follows. When the correspondence to image content is considered in indexing videos, it often happens that the organizational name indicates a particular place in the image ("XX bank," for example, indicates the image of a bank building, as well as the organization occupying it), so we do not consider a clear distinction between places and organizations in this work.

In the following, existing studies related to the analysis of nouns (noun phrases) that have some use for the purpose of this study are discussed. Then, the semantic attribute analysis method adapted to the nature of captions which is used in this study is described.

2.2.1. Related studies

An example of a similar semantic attribute analysis problem is the named entity task [16, 20], which is defined as a task in the Message Understanding Conference (MUC) for English, and the Information Retrieval and Extraction Exercise (IREX) for Japanese. In this problem, a task is posed in which the participant is asked to extract representations concerning persons, places/organizations, times, numbers, and so on (the first three are limited to proper nouns) from presented text, and to annotate them.

In contrast to these approaches, there is an approach to analyze the semantics of a proper noun by syntax analysis of general text [14]. These problems and methods are close to the technique required in this study, in the sense that the expressions concerning persons and places/organizations are extracted and classified. They differ, however, in that the object of analysis is limited to proper expressions, especially proper nouns. An analysis for only proper nouns and an analysis that includes general noun phases differ fundamentally in their difficulties. Without auxiliary information, it is difficult to determine the semantics of a proper noun such as "mori" (forest, see later discussions) as an isolated word. The semantics must be inferred from the case structure of the sentence and the context of neighboring sentences, as in the method of Nasukawa [14] and various other methods of proper expression extraction.

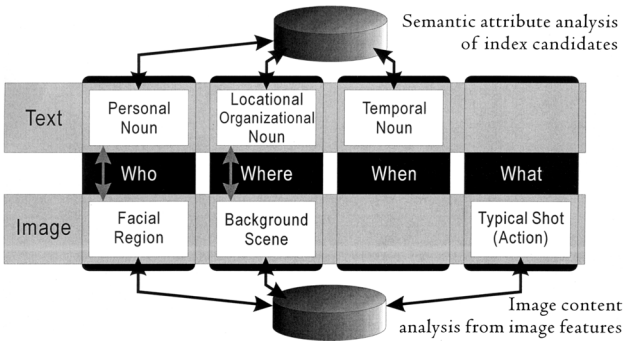

Fig. 1. Indexing considering correspondences between indices and image contents.

Watanabe and colleagues proposed a method of caption analysis for news videos that takes account of the location of the caption and its grammatical features [19]. However, this method has the problem that the designing policy, including the position in the image, differs from program to program, which reduces the versatility of the method.

2.2.2. Analysis based on suffix nouns

The above properties of caption and problems of existing methods were considered, and a method was adopted in which semantic attributes are analyzed for the caption alone, with no possibility of referring to adjacent context and case structure.

It is in general very difficult even for a human to decide whether a particular noun is a common noun or a proper noun. It is also difficult to decide, in the case of a proper noun, whether it is a personal name, locational name, or organizational name from that noun alone, without considering the context. In practice, the semantics of stems or of whole noun phrases in Japanese seems to be governed by the tail noun (suffix noun) as shown in Table 1.

In the above example, it is difficult to determine the semantics from just the word "mori" (forest). However, if it is connected to "prime minister," it is seen to be a personal name, and if it is connected to "town," it is seen to be a locational name. Thus, in Japanese in general, the semantics of whole noun phrases can be determined by analyzing the suffix noun. Studies based on such an assumption have been applied to the recognition of nouns, including proper nouns [7, 13]. In this study such a property is assumed, and the semantic attributes of news captions composed of noun phrases are analyzed by focusing on the suffix nouns.

Figure 2 shows an example of semantic attribute analysis of a caption based on this principle. First, morphological analysis is applied to a caption, and the suffix noun is extracted. Then the extracted suffix noun is analyzed by comparing it to the words in attribute-tagged suffix noun dictionaries.

In this example and the following discussion the suffix noun is emphasized, but the proposed method essentially considers and analyzes not only noun phrases including proper nouns, but also the nouns, such as "actor" and "kitchen," which by themselves indicate a person or a place.

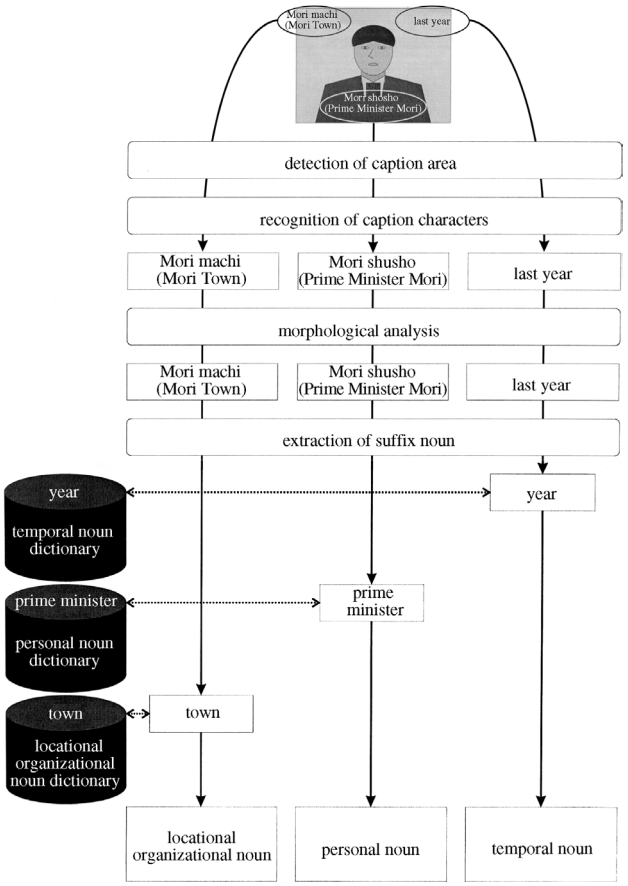

Fig. 2. Example of caption analysis referring to suffixes.

3. Compilation of Semantic Attribute-Tagged Suffix Noun Dictionaries

This section discusses the construction of the semantic attribute-tagged suffix noun dictionaries which are needed in performing semantic attribute analysis of new captions focusing on the suffix noun, as described in the previous section. More precisely, separate dictionaries are constructed for suffix nouns representing (1) persons, (2) locations/organizations, and (3) time.

As already discussed, there are nouns such as "actor," "kitchen," and "today," which by themselves represent persons, locations/organizations, and time, in addition to pure suffix nouns. These nouns are also acquired. Proper nouns are excluded from the acquisition process, since it is difficult to determine their semantics from the words in isolation.

In the compilation of the dictionary, two text corpora were used. These text corpora are tagged with the results of manual morphological analysis. The segmentation and classification of morphemes were assumed to be correct, and were utilized in the study:

- RWC-DB-TEXT-95-2 in the RWC Text Database (2nd version) [2]. This consists of 27,418 sentences from 1994 issues of the Mainichi Shimbun newspaper.
- The Kyoto University Text Corpus (2nd version) [11]. This consists of 19,956 sentences from 1995 issues of the Mainichi Shimbun newspaper.

These two corpora were combined, to comprise 47,374 sentences, and the suffix nouns were extracted under certain conditions. Below, the compilation procedure for each semantic attribute-tagged suffix noun dictionary is described, together with the results of acquisition.

3.1. Acquisition of personal nouns

This section considers the personal nouns, that is, common nouns with a suffix noun indicating a person, such as "XX hakase (Dr. XX)," as well as common nouns that themselves indicate a person, such as "actor." The acquisition procedure and the results are described.

3.1.1. Extraction from the corpus

Personal nouns were extracted from the corpus according to the criteria shown in Table 2. We took advantage of the property that the suffix auxiliaries "(ra)" and "(tachi)" are usually applied in Japanese to represent (exclusively) multiple persons. The examples of extraction by this criterion are shown in the bottom row of Table 2. Table 3 lists some of the extracted words in decreasing order of frequency.

Table 1. Example of determination of stem's meaning by the suffix

Mori	...	Mori	[noun–?]		
Mori + "shusho" (prime minister)	...	Mori	[noun–proper noun–personal name]	+ "shusho" (prime minister)	[noun–common noun]
Mori + "machi" (town)	...	Mori	[noun–proper noun–locational name]	+ "machi" (town)	[noun–common noun]

Table 2. Personal noun extraction rules: Extract the underlined word from a sequence that matches the following patterns of both words and morphemes. "*" denotes that any word or morpheme is applicable. Patterns for both RWC corpus and Kyoto University corpus are shown, since they have different morphological systems. The extraction example shows the result from automatic morphological analysis performed by the Japanese morphological analysis system JUMAN.

Word	*	...	*	["ra" (plural suffix)] ["tachi" (plural suffix)]
RWC morpheme	[noun–*]	...	[noun] [noun–nonautonomous–*] [noun–suffix]	[noun–suffix]
Kyoto Univ. morpheme	[noun–*] [suffix–noun type suffix noun]	...	[noun–common noun] [suffix–noun type suffix noun]	[suffix–noun type suffix noun]
Example	Tanaka [noun–personal name]	Ichiro [noun–personal name]	"kyoju" (professor) [noun–common noun]	"ra" (plural suffix) [suffix–noun type suffix noun]

3.1.2. Expansion of vocabulary by use of a thesaurus

By the criteria shown in Table 2, 136 suffix nouns were extracted from the corpus. However, the vocabulary was still insufficient for use in actual news caption analysis. Consequently, the vocabulary was expanded by using a thesaurus. In this expansion, a word belonging to the same class as the extracted word was always judged to belong to the semantic class in question.

As the thesaurus we used the "lexical class table in the classified lexical table format (revised edition)" [8] (below called the "new classified lexical table"). It is a concept class dictionary in which 87,743 words (70,858 distinct words) are assigned to 4 main classes, 842 subclasses, and 10,334 class items. The class item, which is the minimum unit of classification, contains 6.8 words on average. Table 4 shows the change of vocabulary size in the course of dictionary compilation. A total of 1795 personal nouns were acquired in the final stage.

Table 3. List of extracted personal nouns (Top 30)

Order	Extracted word	Occurrences	Order	Extracted word	Occurrences
1	"sha" (person)	75	14	"shokicho" (secretary general)	8
2	"shi" (Mr./Ms.)	57	14	"kyaku" (guest)	8
3	"san" (Mr./Ms.)	51	14	"in" (member)	8
4	"kaicho" (president)	32	19	"daihyo" (representative)	7
5	"cho" (head)	26	20	"yogisha" (suspected)	6
6	"giin" (representative of House)	21	20	"shokuin" (clerk)	6
7	"nin" (person)	16	20	"jokyoju" (associate professor)	6
7	"kyoju" (professor)	16	20	"kyokucho" (head of department)	6
9	"ka" (specialist)	15	20	"kankeisha" (related person)	6
10	"kanbu" (key person)	14	20	"kan" (clerk)	6
11	"shusho" (prime minister)	13	20	"iincho" (chairman)	6
12	"hikoku" (accused)	10	27	"taishi" (ambassador)	5
13	"gyosha" (contractor)	9	27	"shacho" (president of company)	5
14	"chokan" (director)	8	27	"kanji" (secretary of party)	5
14	"sho" (minister)	8	27	"kacho" (head of section)	5

Table 4. Number of collected personal nouns: Numbers of classes and subclasses indicate those that extracted words listed in the thesaurus belong to

	Extraction from corpus	Expansion by New Classified Lexical Table
Word in New Classified Lexical Table	117	1,776
(number of classes)	125	125
(number of subclasses)	59	59
Words outside of New Classified Lexical Table	19	19
Total thesaurus	136	1,795

3.2. Acquisition of locational/organizational nouns

Locational/organizational nouns—suffix nouns indicating a location or organization, such as "XX station"—and common nouns which by themselves indicate a location or organization, such as "kitchen," were acquired. This section describes the acquisition procedure and the results.

3.2.1. Extraction from corpus

Locational/organizational nouns were extracted from the corpus on the basis of the criteria shown in Table 5. In this process, the fact that the postpositional case auxiliaries "kara" (from), "de" (at), "ni" (to), "e" (to), "yori" (from), and "nite" (at) are used to indicate location or direction was utilized. Examples of extraction by this criterion are given in the second row from the bottom in Table 5.

3.2.2. Elimination of personal nouns

The criteria shown in Table 5 are not rigorous, and words which are essentially personal nouns were included in the extraction based on these criteria, as shown in the bottom row. This is due to the multiple usage of the postpositional auxiliaries. The postpositional case auxiliary "e" (to), for example, indicates direction in the extraction described in this section, but can also indicate a person in the dative case in the included personal nouns. The included personal nouns were eliminated by removing the personal nouns extracted in Section 3.1 from among the words extracted in this section. Table 6 lists some of the extracted

Table 5. Locational/organizational noun extraction rules

Word	*	*	...	*	["kara" (from)] ["de" (at)] ["ni" (to)] ["e" (to)] ["yori" (from)] ["nite" (at)]
RWC morpheme	[noun–proper noun] [noun–proper noun–organization] [noun–proper noun–location–*]	[noun–*]	...	[noun] [noun–non-autonomous–*] [noun–suffix]	[postpositional auxiliary–case auxiliary]
Kyoto Univ. morpheme	[noun–locational name] [noun–organizational name]	[noun–*] [suffix–noun type suffix noun]	...	[noun–common noun] [suffix–noun type suffix noun]	[postpositional auxiliary–case auxiliary]
Example	"Tokyo" [noun–locational name]	"daigaku" (university) [noun–common noun]	"kogakubu" (faculty of engineering) [noun–common noun]	"konai" (within campus) [noun–common noun]	"e" (to) [postpositional auxiliary–case auxiliary]
Example of mistakenly mixed personal name	"Tokyo" [noun–locational name]	"daigaku" (university) [noun–common noun]	"kogakubu" (faculty of engineering) [noun–common noun]	"kyoju" (professor) [noun–common noun]	"e" (to) [postpositional auxiliary–case auxiliary]

words after elimination of personal nouns, in decreasing order of frequency.

3.2.3. Expansion of vocabulary by use of a thesaurus

A total of 696 suffix nouns were left after eliminating personal nouns. This vocabulary was insufficient for use in actual caption analysis. Consequently, the vocabulary was expanded by using the thesaurus as described in Section 3.1.

Table 7 shows the change in the vocabulary size during the construction of the dictionary. Some 7908 locational/organizational nouns were acquired in the final stage.

3.3. Acquisition of temporal nouns

Temporal nouns—suffix nouns indicating time, such as "XX go" (after XX)—and common nouns indicating time by themselves, such as "today," were acquired. This section describes the acquisition procedure and the results.

3.3.1. Extraction from corpus

Temporal nouns were extracted from the corpus based on the criteria shown in Table 8. The property of the postpositional case auxiliaries that "kara" (from), "ni" (to), and "yori" (since) are used to indicate time was utilized. Although not shown in Table 8, the "noun–temporal noun" morphemes were prespecified in the Kyoto University Text Corpus, and all words with those morphemes were extracted. An extraction example based on this criterion is shown in the bottom row of Table 8. Table 9 lists some of the extracted words in decreasing order of frequency.

Table 7. Number of collected locational/organizational nouns

	Extraction from corpus	After deleting personal name	Expansion by New Classified Lexical Table
Word in New Classified Lexical Table	674	607	7,819
(number of classes)	764	697	697
(number of subclasses)	318	307	307
Words outside of New Classified Lexical Table	91	89	89
Total thesaurus	765	696	7,908

3.3.2. Expansion of vocabulary by use of a thesaurus

Some 156 suffix nouns were extracted from the corpus based on the criteria shown in Table 8. However, this number is still insufficient for actual caption analysis. Consequently, the vocabulary size was expanded by using the thesaurus as described in Section 3.1. The change in the

Table 6. List of extracted locational/organizational nouns (Top 30)

Order	Extracted word	Occurrence	Order	Extracted word	Occurrence
1	"shi" (city)	163	16	"kaidan" (conference)	28
2	"nai" (in)	108	16	"sen" (election)	28
3	"gawa" (side)	78	18	"chisai" (regional court)	27
4	"taikai" (convention)	60	19	"machi" (town)	25
5	"ken" (prefecture)	46	20	"ichiba" (market)	23
6	"shinai" (in a city)	43	20	"kan" (between)	23
7	"eki" (station)	39	22	"chiku" (region)	22
7	"seifu" (government)	39	22	"kai" (party)	22
7	"sen" (war)	39	24	"kan" (building)	20
10	"mondai" (problem)	38	25	"oki" (offshore)	19
11	"sho" (department)	36	25	"kaigi" (conference)	19
12	"chiho" (district)	31	25	"chiiki" (region)	19
13	"ku" (ward)	30	28	"mura" (village)	18
14	"daishinsai" (catastrophic earthquake)	29	28	"ba" (site)	18
14	"kokunai" (in a country)	29	28	"gun" (troops)	18

Table 8. Temporal noun extraction rules

Word	*	["kara" (from)] ["ni" (to)] ["yori" (from)]
RWC morpheme	[noun–adverb–*] [noun–non-autonomous–adverb possible–*]	[postpositional auxiliary–case auxiliary]
Kyoto Univ. morpheme	[noun–adverb type noun]	[postpositional auxiliary–case auxiliary]
Example	"koro" (about) [noun–adverb type noun]	"kara" (from) [postpositional auxiliary–case auxiliary]

vocabulary size in the course of dictionary compilation is shown in Table 10. A total of 2144 temporal nouns were acquired at the final stage.

4. Evaluation of Dictionary Performance by Analysis of News Captions

Using the semantic attribute-tagged suffix noun dictionaries compiled as described above, actual news video captions were analyzed and the performance of the dictionaries evaluated.

Table 10. Number of collected temporal nouns

	Extraction from corpus	Expansion by New Classified Lexical Table
Word in New Classified Lexical Table	136	2,124
(number of classes)	213	213
(number of subclasses)	104	104
Words outside of New Classified Lexical Table	20	20
Total thesaurus	156	2,144

4.1. Conditions of experiment

The captions were analyzed by the procedure shown in Fig. 2. In this study, the caption area was detected and the characters were recognized by human observation and the results were transcribed. For the morphological analysis, the Japanese morpheme analysis system JUMAN [10] was used. JUMAN outputs the borders between morphemes and their attributes/parts of speech. In this study, if the tail of the caption to be analyzed was a noun, the caption

Table 9. List of extracted temporal nouns (Top 30)

Order	Extracted word	Occurrence	Order	Extracted word	Occurrence
1	"tame" (for)	465	16	"hoka" (other)	30
2	"jidai" (era)	145	17	"koro" (about)	27
3	"chu" (in)	128	18	"sengo" (after war)	25
4	"kan" (between)	123	19	"toji" (at that time)	24
5	"zen" (before)	117	19	"jiki" (time)	24
6	"ji" (time)	116	19	"mama" (remain)	24
7	"tokoro" (place)	79	22	"toki" (when)	23
8	"uchi" (within)	65	23	"kekka" (result)	22
9	"sai" (when)	54	24	"kotoshi" (this year)	20
10	"baai" (case)	47	24	"genzai" (at present)	20
11	"ue" (above)	46	26	"tabi" (time)	16
12	"chokugo" (immediately after)	45	26	"chokuzen" (immediately before)	16
13	"irai" (since)	43	26	"choki" (long time)	16
14	"nichi" (day)	40	29	"iko" (after)	15
15	"go" (after)	31	29	"naka" (in)	15

was compared to the semantic attribute-tagged suffix noun dictionaries and was analyzed.

JUMAN itself contains a proper noun dictionary and a temporal noun dictionary. For captions with a personal name, locational name, or organizational name at the tail, or some temporal nouns, the output result of JUMAN was given priority for use. Only nouns which could not be judged to be proper nouns or temporal nouns were analyzed by the proposed method. Figure 3 shows the method of utilizing the dictionaries.

A total of 2546 captions appearing in approximately 370 minutes of news videos were used in the experiment. Captions other than noun phrases, such as headline title, as well as captions appearing in schematic diagrams using CG or flips, were excluded from the experiment. The reason is that these captions rarely describe the contents of the image directly. These captions should be easily identified by image recognition and natural language processing.

The correct solution in the evaluation is provided by a third party. When the result of analysis indicated more than one possible attribute, the result was judged to be correct if the correct result was included.

4.2. Experimental results and discussion

The experimental results obtained under the above conditions are now presented and discussed.

4.2.1. Experimental results

The results of the analysis are shown in Tables 11 to 13. In the tables, the numbers under "analysis by proposed method" are the results of analysis by the method shown in Fig. 3. The numbers under "analysis by JUMAN" are the results of analysis by JUMAN. As already described, captions with a pronoun, which is a personal name or a locational or organizational name at the tail, as well as some temporal nouns, were analyzed by JUMAN. "Combined" denotes the results of analysis by combining the two procedures.

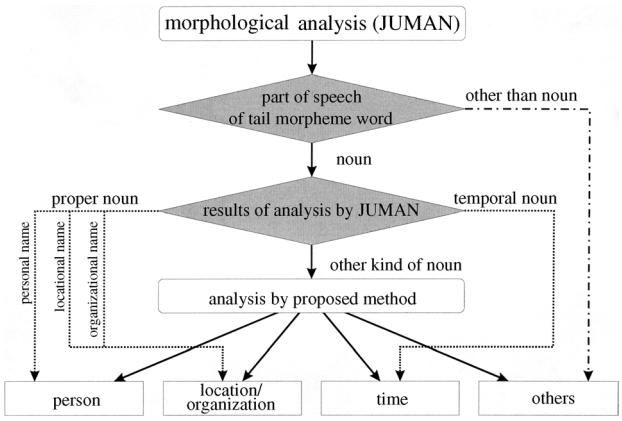

Fig. 3. Analysis process by combination of JUMAN and the proposed method.

The following definitions are used. The precision and the recall were used as evaluation measures. The precision in each result expresses the performance by that analytical procedure alone. The recall expresses the contribution of each analytical procedure to the overall recall.

$$\text{precision} = \frac{\text{number of correct results } (N_c)}{\text{number of correct results } (N_c) + \text{number of mistakes } (N_m)}$$

$$\text{recall} = \frac{\text{number of correct results } (N_c)}{\text{number of correct results } (N_c) + \text{number of oversights } (N_o)}$$

4.2.2. Discussion

In the experiments, high recall was obtained for each semantic attribute. This is an important factor in deriving the index for particular semantics, constrained by other processing operations such as image processing. The precision is generally low, both in the proposed method alone and in the combined method, except for captions indicating a person. The reasons are that inadequate words existed in the compiled dictionary, and that the proposed method is not sufficient to assure correct analysis, due to semantic diversity.

In the analysis of captions indicating a person, a recall and precision that were almost entirely satisfactory for practical purposes were obtained. Although the precision of the proposed method is lower than the results of IREX, which similarly handled Japanese text [21], the recall is almost equal. The accuracy is sufficient for practical purposes in an application to index videos in which a larger amount of information is collected, with the expectation that the excess information will be selected in later image processing and integration processing.

JUMAN can analyze only the attributes of proper nouns and some temporal nouns. By combining JUMAN with the proposed method, however, noun phrases with a

Table 11. Results of personal caption analysis by the created dictionary: N_c, N_m, and N_o stand for numbers of correct, mistaken, and oversight answers, respectively

	N_c	N_m	N_o	Precision	Recall
Analysis by proposed method	210	54	–	79.6%	68.9%
Analysis by JUMAN	33	45	–	42.3%	10.8%
Combined	243	99	62	**71.1%**	**79.7%**

Table 12 Results of locational/organizational caption analysis by the created dictionary

	N_c	N_m	N_o	Precision	Recall
Analysis by proposed method	307	491	–	38.5%	46.9%
Analysis by JUMAN	220	12	–	94.8%	33.6%
Combined	527	503	127	**51.3%**	**80.5%**

common noun as the tail noun can be analyzed regardless of whether or not a proper noun is included. Thus, the analysis performance was improved to 46 to 83% in terms of recall.

The main reasons for errors and oversights were the five shown in Table 14. Items (a) and (e) are direct reflections of the performance of the dictionaries that were compiled. Approximately 40% of the cases attributed to item (a) should have been analyzed correctly if the sentence was decomposed into more detailed morphemes. One method of handling such a situation may be to use the criterion of longest late fit, not complete fit as is used at present. However, this approach may increase mistakes.

Item (e) is a composite problem due to the existence of inadequate words. It results from the less rigorous extraction conditions shown in Tables 2, 5, and 8, as well as the presence of inappropriate words that were included while expanding the vocabulary by using the thesaurus. This problem can be resolved by manual correction of the dictionaries. Most of the inadequate words in the former problem were extracted due to the use of postpositional auxiliaries in a manner different from the use assumed in the extraction conditions. In order to resolve this problem, the morpheme information in the corpus should include the use of postpositional auxiliaries. Most of the inappropriate words associated with the latter problem are due to discrepancies between the class specified by the classification principle in the thesaurus that was employed and the class used in the proposed method. Consequently, improvement should result from using a dictionary with a classification principle geared to the proposed method.

Table 13. Results of temporal caption analysis by the created dictionary

	N_c	N_m	N_o	Precision	Recall
Analysis by proposed method	142	221	–	39.1%	83.5%
Analysis by JUMAN	17	0	–	100.0%	10.0%
Combined	159	221	11	**41.8%**	**93.5%**

Table 14. Reasons for mistakes and oversights

	Reason	Ratio
(a)	Insufficient vocabulary for common noun in the compiled dictionaries	44%
(b)	Insufficient vocabulary in JUMAN proper noun and temporal noun dictionaries	39%
(c)	Diversity of semantics (example) “Konishiki-zeki” (name of a *Sumo* wrestler, person), “Hakone-seki” (Hakone checkpoint, place/organization)	9%
(d)	Incorrect morpheme analysis by JUMAN (example) ○ “bnen-sei” (6th grader, person), X “b-nensei” (others)	6%
(e)	Existence of inadequate words in the compiled dictionaries	2%

Item (c) is an essential problem which arises in handling semantics. It cannot be easily resolved by a simple analysis as used in this study. To deal with items (b) and (d), there must be improvements in the vocabulary of the morphological analysis tool and improvement of analytical performance.

Table 15 compares the results obtained by using dictionaries (ideal dictionaries) in which adequate classification items were manually selected from the thesaurus, and the results shown in Tables 11 to 13. This is a comparison of the upper bound of semantic attribute analysis by the proposed method. Since the number of correct results is increased only slightly if an ideal dictionary is used, we see that item (a) is not a serious problem in terms of the number of occurrences. On the other hand, the number of mistakes is greatly reduced, which indicates that item (e) is a serious problem in terms of the number of occurrences. Since the incidence of oversights (N_o) is not reduced much if the ideal dictionary is used, it is possible that the vocabulary not included in the thesaurus that we employed should be complemented in order to fill gaps. That is, additional words should be extracted from a larger-scale corpus by using the proposed method.

5. Conclusions

This paper has described the process of composing a dictionary for the semantic attribute analysis of captions, to

Table 15. Results of caption analysis by manually created dictionary

	N_c	N_m	N_o	Precision	Recall
Person	243 (±0)	93 (−6)	62 (±0)	72.3% (+1.2%)	79.7% (±0.0%)
Place/ organization	534 (+7)	172 (−331)	120 (−7)	75.6% (+24.3%)	81.7% (+1.2%)
Time	159 (±0)	29 (−192)	11 (±0)	84.6% (+42.8%)	93.5% (±0.0%)

be used in automatic indexing of news videos. A performance evaluation has been presented. Specifically, the noun (suffix noun) at the tail of the caption (noun phrase) is noted. To support an analysis of classifications of personal, locational/organizational, and temporal uses, the dictionaries were constructed by extracting suffix nouns indicating semantic attributes from corpora.

The resulting dictionaries consisted of 1795 words in the personal noun dictionary, 7908 words in the locational/organizational noun dictionary, and 2144 words in the temporal noun dictionary, for a total of 11,847 words. It is conceivable that a similar dictionary could be constructed by manually selecting the classification items in the New Classified Lexical Table. However, automatic composition saves the effort of manual composition and yields the supporting sample data. Another advantage is that words not contained in the New Classified Lexical Table can be included. In this study, 19 personal words, 89 locational/organizational words, and 20 temporal words, or a total of 128 words, not included in the table were acquired.

As the next step, the proposed method was applied to the analysis of actual news videos and its performance was evaluated. Our analysis showed that the recall, which is important in indexing, was high, indicating the practical usefulness of the method, although the precision was low. Comparison experiments using an ideal dictionary failed to remedy the problem of insufficient vocabulary of the dictionary. It will be necessary to add words not contained in the thesaurus in order to improve the recall, which is important in indexing. The above result suggests that it will be necessary to enlarge the range of word acquisition from a larger-scale corpus.

For the contents of the dictionary that was constructed (classification items and words outside the classified lexical table), see Ref. 5. Regarding the results of indexing actual news videos by using the dictionary compiled in this study and the mechanism shown in Fig. 1, see Ref. 4.

Acknowledgments. The "Lexical Classification Table in the 'Classified Lexical Table' Format (Revised Version)" [8] was used under a monitor contract with the National Institute of Japanese Language. The "RWC Text Database" [2] was used under a license contract with the Real World Computing Project (RWCP). A tremendous amount of labor was contributed by Mr. H. Taira and Ms. T. Oda in transcribing the caption texts to be used in the experiment, for which the authors are grateful.

REFERENCES

1. Ando A, Imai T, Kobayashi A, Homma S, Goto A, Kiyoyama N, Mishima T, Kobayakawa T, Sato S, Onoe K, Segi H, Imai A, Matsui A, Nakamura A, Tanaka H, Togi T, Miyasaka E, Isono H. A broadcast news caption composition system based on speech recognition. Trans IEICE 2001;J84-D-II:877–887.
2. Real World Computing Project (RWCP). RWC Text Database, 2nd version, 1998.
3. Hori O, Mita Y. A robust method of character recognition from video for telop recognition. Trans IEICE 2001;J84-D-II:1800–1808.
4. Ide I, Hamada R, Sakai S, Tanaka H. An attribute based news video indexing. Proc ACM Multimedia 2001 Workshops—Multimedia Information Retrieval, p 70–73.
5. Ide I. A study on automatic indexing of video. Indexing by integrated media processing and its application to news video. Doctoral dissertation, Graduate School of Engineering (Electrical Engineering), The University of Tokyo, 1999.
6. Ide I, Yamamoto K, Hamada R, Tanaka H. Automatic indexing of videos based on shot classification. Trans IEICE 1999;J82-D-II:1543–1551.
7. Kato N, Uraya N, Aizawa T, Nakase S. Processing of proper nouns in English–Japanese translation. 40th Natl Conv Inf Process Soc 1990;2F-2, p 421–422.
8. National Institute of Japanese Language Lexical Classification Table in 'Classified Lexical Table' Format (Revised Version), Revised Electronic Data for Monitoring, 1996.
9. Kurakake S, Kuwano H, Odaka K. Recognition and visual feature matching of text region in video for conceptual indexing. Proc SPIE Conf 3022: Storage and Retrieval for Image and Video Databases V, p 368–379, 1997.
10. Artificial Intelligence Language Media Laboratory, Graduate School of Information Science, Kyoto University. Japanese Morphological analysis system JUMAN version 3.6. http://www-lab25.kuee. kyoto-u. ac.jp/nl-resource/juman.html, 1998.

11. Artificial Intelligence Language Media Laboratory, Graduate School of Information Science, Kyoto University. Kyoto University Corpus, version 2.0. http://www-lab25.kuee.kyoto-u.ac.jp/nl-resource/corpus.html, 1998.
12. Nakamura Y, Kanade T. Semantic analysis for video contents extraction—Spotting by association in news video. Proc Fifth ACM Int Multimedia Conf (ACM Multimedia '97), p 393–402.
13. Nagase T. Pre-processing for Japanese syntax analysis based on morpheme type. 41st Natl Conv Inf Process Soc 1990, 1S-6, Vol. 3, 109–110.
14. Nasukawa T. Determination of keyword semantics based on context information. 11th Natl Conv Soc Artif Intell 1997; 17-1, p 348–349.
15. Satoh S, Nakamura Y, Kanade T. Name-It: Naming and detecting faces in news videos. IEEE MultiMedia 1999;6:22–35.
16. Sundheim BM. Named entity task definition, version 2.1. Proc Sixth Message Understanding Conf (MUC-6), 1995, p 317–332.
17. Wactler HD, Hauptmann AG, Witbrock MJ. Informedia News-on-Demand: Using speech recognition to create a digital video library. CMU Tech Rep CMU-CS-98-109, Carnegie Mellon University, 1998.
18. Wactler HD, Christel MG, Gong Y, Hauptmann AG. Lessons learned from building a terabyte digital video library. IEEE Comput 1999;32:66–73.
19. Watanabe Y, Okada N, Nagao M. Semantic analysis of telops used in video news. Tech Rep Nat Lang Process Inf Process Soc 1996, 96-NL-116, p 107–114.
20. Appendix H: NE definition. Proc IREX Workshop 1999, p 264–273.
21. Appendix J: Results of NE evaluation. Proc IREX Workshop 1999, p 286–294.

AUTHORS (from left to right)

Ichiro Ide (member) received his B.S. degree from the Department of Electronics Engineering, The University of Tokyo, in 1994, completed the M.E. program (information engineering) and doctoral program (electrical engineering) in 1996 and 2000, and joined the National Institute of Informatics as a research associate. His research interests are natural language processing, video understanding, and integrated media processing. He received a 1995 Encouragement Award from the 51st National Convention of the Information Processing Society of Japan. He holds a D.Eng. degree, and is a member of the Japanese Society for Artificial Intelligence and the Information Processing Society of Japan.

Reiko Hamada (student member) received her B.S. degree from the Department of Information and Communication Engineering, The University of Tokyo, in 1998, completed the M.E. program (electrical engineering) in 2000, and is now in the doctoral program. She is a JSPS Special Researcher. She received a 2000 Encouragement Award from the 63rd National Convention of the Information Processing Society of Japan. She is a student member of the Information Processing Society of Japan.

AUTHORS (continued) (from left to right)

Shuichi Sakai (member) received his B.S. degree from the Department of Information Science, The University of Tokyo, in 1981, completed his graduate course in information engineering in 1986, and joined the Electro-technical Laboratory. He was an invited researcher at MIT in 1991–92; head of the RWC Super-Parallel Architecture Laboratory, 1993–94; associate professor, Institute of Electronics and Information Engineering, University of Tsukuba, 1996–98; associate professor, Graduate School of Engineering, The University of Tokyo, 1998; professor, Graduate School of Information Science and Engineering, 2001. He is engaged in research on computer systems in general, especially architecture, parallel processing, scheduling problems, and multimedia. He received a 1990 Information Processing Society of Japan Best Paper Award, 1991 IBM Japan Award, 1995 Ichimura Science Award, and 1995 71CCD Outstanding Paper Award. He holds a D.Eng. degree, and is a member of the Information Processing Society of Japan and the Japanese Society for Artificial Intelligence.

Hidehiko Tanaka (member) received his B.S. degree from the Department of Electronics Engineering, The University of Tokyo, in 1965, completed the doctoral program (electrical engineering) in 1970, and became a lecturer there. He became a professor on the Faculty of Engineering in 1987, and is now a professor in the Graduate School of Information Science and Engineering. He was a visiting professor at New York City University in 1978–79. His research interests are computer architecture, parallel processing, artificial intelligence, media processing, natural language processing, distributed processing, and CAD. He is the author of *Non-Neumann Computers* and *Information Communication Systems*; a coauthor of *Computer Architectures*, *VLSI Computing I, II*, and *Software-Oriented Architecture*; and Editor of *New Generation Computing*. He holds a D.Eng. degree, and is a member of the Information Processing Society of Japan, the Japanese Society for Artificial Intelligence, and the Japan Society for Software Science and Technology.