Semantic Analysis of Television News Captions Referring to Suffixes

Ichiro IDE, Reiko HAMADA, Shuichi SAKAI, and Hidehiko TANAKA
Graduate School of Electrical Engineering, The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
TEL: +81-3-5841-7413, FAX: +81-3-5800-6922
{ide,reiko,sakai,tanaka}@mtl.t.u-tokyo.ac.jp

Abstract

Semantic analysis to keyword candidates is essential for high quality automatic indexing to video. We propose an improved caption (*i.e.* noun phrase) analysis method for television news video indexing. The analysis is performed by taking advantage of the characteristics of Japanese noun phrases, that the last noun (suffix) determines the semantics of the entire phrase. In this paper, we will present the collection process of such suffixes, and the evaluation of the dictionaries by analyzing actual television news captions. Although the evaluation did not show high precision, it showed 82% to 93% recall, which is important when finding index candidates with specific semantic attributes.

1 Introduction

As the amount of broadcast video data increases, it is becoming more and more important to store them in a well organized manner, considering their recycling and retrieval. Above all, television news programs are worthwhile indexing considering the importance and usefulness. Currently this process is mostly done manually, but automatic indexing is in big demand both to cope with the increasing amount and to achieve sufficient precision for detailed retrieval.

We are trying to accomplish this task by referring to both video data and accompanying textual data of television news programs. In order to enable high quality indexing for detailed retrieval, simple tagging of keywords as seen in many conventional methods are insufficient, and semantic analysis of keyword candidates is indispensable.

There have been various attempts to automatically index television news video from this approach as prominent in the Informedia project's News-on-Demand system [Wactler98]. Nonetheless, although they do satisfy the demand for automatic indexing to a certain extent, their indexing strategies are mostly based on statistics, or just simple occurrence of words and phrases. These strategies do not necessarily ensure the correspondence between the indexes and the image contents. Although this may satisfy demands for retrieving a whole news topic, it is not sufficient for retrieving short video segments, where keyword candidates may exist in adjacent segments. Moreover, ensuring such correspondence is crucial, in order to provide indexes that depict certain topics occurring in such short segments.

Reflecting this issue, we are currently working on an automatic video indexing system, which performs indexing considering such correspondences. In the natural language processing part of this system, semantic analysis of keyword candidates is required. This is a common technology required for other similar indexing methods [Ide99, Satoh99, Nakamura97]. This paper describes and evaluates a news caption analysis method for such purpose. Concretely, the semantic analysis mentioned here is identical to classification of noun phrases into four different classes; namely, (1) personal, (2) locational / organizational, (3) temporal, and (4) others.

Note that we have presented the basic idea of the proposed analysis method in [Ide98], but this paper presents the result of refined classification dictionary creation, and evaluation performed to larger amount of data.

Although we will concentrate on news caption analysis in this paper, the collected dictionary should be useful for applying to such tasks as the *Named Entity task* [NEtask], and for extracting and classifying personal, locational, and organizational proper noun for further natural language context analysis purposes.

Section 2 introduces the analysis method, and Section 3 describes the collection of suffixes that are used for the analysis. Section 4 shows the process and result of an evaluation of the collected dictionary applied to actual television news captions, and Section 5 concludes the paper.

2 News Caption Analysis Referring to Suffixes

In this Section, we will first present characteristics of television news captions, and then propose our semantic analysis method.

There are various textual data accompanying a video other than (open) caption: main audio, sub audio and closed caption. Although they contain much information, it is generally considered difficult to extract semantically important sentences and keywords from these sources. Moreover, although accuracy of speech recognition is improving, their results are still not sufficiently reliable as a source of index candidates. Closed caption may resolve this problem, but since Japanese character input method requires transformation of syllabic characters (*Kana*) to Chinese origin characters (*Kanji*), it is still not realistic to expect them to be provided with live broadcast news videos.

On the other hand, (open) captions usually indicate important information, which frees us from extracting index candidates from other sources. Character recognition for caption has been challenged by many groups [Hori99, Sawaki98]. Since captions overlap background images, and characters tend to be in low resolution in proportion to the few number of scanning lines (525 in the case of NTSC broadcasting standard), it is difficult to apply conventional OCR (Optical Character Recognition) technologies. Nonetheless, their works have achieved certain progress.

Thus, we employ captions for keyword sources, and will discuss about their analysis in this paper. Note that currently caption recognition is performed manually, and will be automated applying some of the above-mentioned technologies in the future.

Table 1: Types and ratios of television news captions: (c) indicates titles in the beginning of a new news topic, (e) includes summaries and translations, and (f) indicates broadcast technical captions such as 'live'.

	Type	Ratio
(a)	Locational / Organizational	28.11%
(b)	Personal	13.19%
(c)	Title	9.43%
(d)	Temporal	7.00%
(e)	Speech	5.52%
(f)	Technical	3.52%
(g)	Descriptive	1.55%
	Others	31.67%

2.1 Characteristics of Captions

Captions have special linguistic characteristics from the point of view of both grammar and semantics. This limits us from employing conventional natural language processing technologies on one hand, but on the other hand we could take advantage of it.

Here grammatical and semantic characteristics are roughly introduced.

2.1.1 Grammatical Characteristics

From the grammatical point of view, captions are peculiar compared with normal texts handled by conventional natural language processing methods. Major characteristics are as follows:

- Sentence-like caption
  1. Simple sentence.
  2. Omission of subject.
  3. Ends with noun.
  4. Ends with specific post-positional particles, such as '*he*', '*ni*', '*ka*', '*mo*', '*ha*', and '*wo*', which indicate estimation, conjecture and so on.
- Non-sentence-like caption
  1. Omission of particles *i.e.* consists mostly of nouns.
  2. Compound noun phrase.

Due to these characteristics, although advanced natural language processing methods do exist, they are not fully applicable to analyze television news captions. More than half of the captions in news video were non-sentence-like, and we will mostly make use of them since they contain important information briefly. This frees us from extracting noun phrases *i.e.* keyword candidates from sentence-like captions.

2.1.2 Semantic Characteristics

On the other hand, captions have a certain trend from the semantic point of view. Table 1 shows the types and their ratios of the captions that appeared in approximately 370 minutes of news video. The table shows that 42.86% ((a),(b), and (g)) could be used as keyword candidates for describing actual objects in the video, and 49.86% ((a), (b), (d), and (g)) for indexing '4W' attributes ('Where', 'Who', 'When', and 'What'). In addition, (c) could be generally used as an rough and abstract index. Including (c), respectively 52.29% and 59.29% of captions could be used as keyword candidates. These useful captions appeared approximately 3.29 and 3.83 times per minute, or 0.42 and 0.49 times per shot, respectively, which is fairly sufficient for indexing.

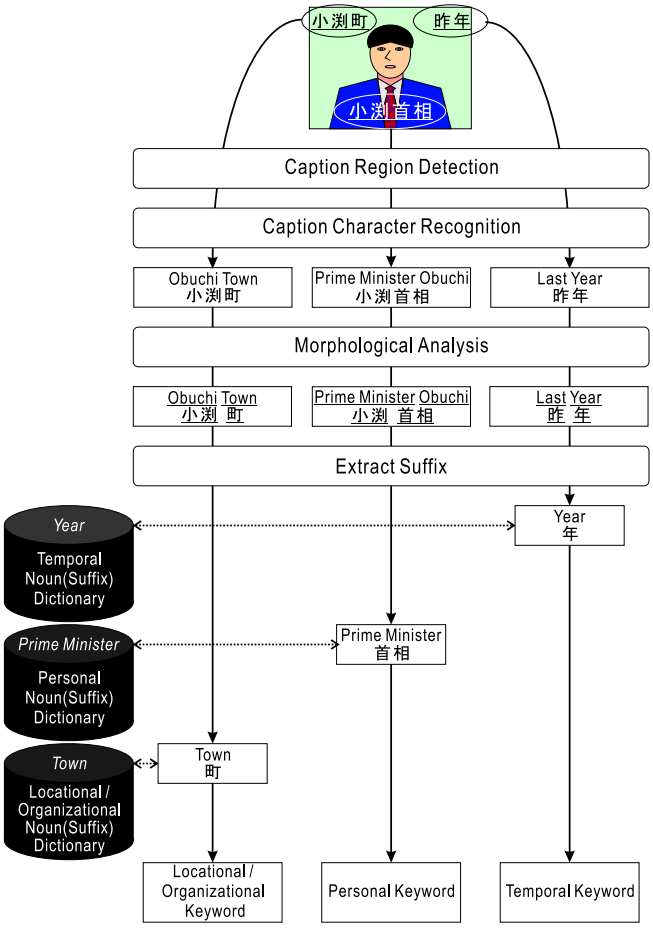

Figure 1: Example of caption analysis referring to suffixes: captions are analyzed according to their suffixes, after morphological analysis. The extracted suffixes are compared to those in each dictionary. Currently, caption region detection and character recognition is done manually.

2.2 Semantic Analysis of Captions

2.2.1 Related Works

As related works, several attempts to analyze nouns that satisfy our purpose to some extent do exist.

Similar to our task, the *Named Entity task* defined for the Message Understanding Conference (MUC) [NEtask] assigns the participants to classify personal, organizational, locational, temporal and numerical phrases.

Nasukawa [Nasukawa97] proposed a semantic disambiguation method that determines semantic attributes of proper nouns (*i.e.* whether a proper noun indicates a place or a person) referring to the context of adjacent sentences.

On the other hand, Watanabe *et al.* [Watanabe96] proposed a method that analyzes television news captions by referring to both location and grammatical characteristics.

Although these tasks are similar to ours, they are difficult to serve our purpose, as follows: The *Named Entity task* limits the tagging to personal, organizational and locational phrases to proper nouns, and Nasukawa's method also is mainly purposed to handle proper nouns. Moreover, captions do not have enough neighboring information to analyze contexts, which makes the application of Nasukawa's method and solutions to the *Named Entity task* difficult to our task. Watanabe *et al.*'s method does serve our purpose, but is not generally applicable to various news programs, which have different designing policies where layouts of captions vary.

2.2.2 Proposed Method: Analysis Based on Suffixes

Considering these issues, we decided to analyze captions on their own. For example, it is impossible even for a human being to distinguish whether '小渕' (Obuchi) alone indicates a person or a place, or even a common noun, which means 'a small pool'. We will generally distinguish them from the suffixes: If '首相' (Prime Minister) follows, as in '小渕-首相' (Prime Minister Obuchi), we know it indicates a person, and if '町' (town) follows, as in '小渕-町' (Obuchi town), we know it indicates a place or an organization.

As shown in this example, it is generally possible in Japanese (and also presumably in other East Asian languages, although the collection rules may somewhat vary) to analyze semantic attributes of noun phrases referring to suffixes. Under this assumption, we will analyze captions that are mostly compound noun phrases, by referring to suffixes.

Although in this example and also in the following Sections, issues on suffixes are mainly discussed, our method also includes analysis to individual nouns such as 'actor' and 'kitchen' that on their own indicate people and places, respectively.

3 Collecting Suffixes from Text Corpora

In this Section, rules and results of collecting suffixes that indicate (1) people, (2) location / organization, and (3) time. Nouns that could individually represent such semantics, such as 'actor', 'kitchen', and 'today' are collected together with genuine suffixes. Proper nouns are excluded from the collection, since they tend to represent both people and location / organization.

The following collection processes were applied to two different text corpora: (1)*RWC-DB-TEXT-95-2* [RWC] and (2)*Kyoto University Corpus* [KUC]. They consist of manually morphological analyzed newspaper articles. Simple descriptions of the two corpora are as follows:

- RWC-DB-TEXT-95-2
  Consists of 27,418 sentences selected from Mainichi newspaper's 1994 edition.
- Kyoto University Corpus
  Consists of 19,956 sentences selected from Mainichi newspaper's 1995 edition.

Consequently, a total of 47,374 sentences were analyzed for collection in the following processes.

3.1 Collecting Personal Suffixes

Here, rules and results of collecting *personal suffixes*; suffixes that indicate people, such as '—, PhD', and individual nouns such as 'actor', are discussed.

3.1.1 Collection from Corpora

Personal suffixes were collected according to the following rule: search for the following pattern, and extract the underlined word.

Common Noun or Suffix
Certain Suffixes

As for the certain suffixes in the collection rule, 'ら (等)' (ra = *et al.*) and 'たち (達)' ($tachi$ = *et al.*) that only follow personal nouns were used.

Following is an example of the collection following this rule. The underlined word is collected as a personal suffix:

総統	(President)	Common Noun
ら	(ra = *et al.*)	Post-positional Particle

Table 2 shows an example of the most frequent words collected by this rule.

3.1.2 Expansion of Vocabulary

Although 136 suffixes were collected, they are still insufficient to apply to general caption analysis. Thus we expanded the vocabulary by referring to the *Classified Lexical Table (Bunrui Goi Hyo) Enlarged Edition Monitoring Version* [CLT][1]. Each sub class that contains at least one of the collected suffixes is considered as appropriate to be included in the dictionary.

Table 3 shows the number of words collected by this process. In the case of words listed in the *Classified Lexical Table*, the numbers of sub classes and classes of which the words belong to are also shown.

3.2 Collecting Locational / Organizational Suffixes

Here, rules and results of collecting *locational / organizational suffixes*; suffixes that indicate location / organization, such as '— station', and individual nouns such as 'kitchen', are discussed.

3.2.1 Collection from Corpora

Locational / organizational suffixes were collected according to the following rule: search for the following pattern, and extract the underlined word.

Locational Name or Organizational Name
Any Noun or Suffix
.....
Common Noun or Suffix
Certain Post-positional Particles

As for the certain post-positional particles in the collection rule, 'から' ($kara$ = from), 'で' (de = at), 'に' (ni = at, to), 'へ' (he = to), 'より' ($yori$ = from), and 'にて' ($nite$ = at) that may indicate location and direction were used.

Following is an example of the collection following this rule. The underlined word is collected as a locational / organizational suffix:

臺北	(Taipei)	Locational Name
市	(city)	Common Noun
南港	(NanKang)	Locational Name
區	(ward)	Common Noun
で	(de = at)	Post-positional Particle

3.2.2 Deletion of Personal Suffixes

Following the rule described in 3.2.1, personal suffixes would also mingle among locational / organizational suffixes, as in the following case:

臺灣	(Taiwan)	Locational Name
総統	(President)	Common Noun
へ	(he = to)	Post-positional Particle

The underlined word is collected as a locational / organizational suffix, which is a mistake. This is due to the diversity of grammatical uses of post-positional particles; in this case, 'へ' (he = to) could either indicate a direction or an objective person. To exclude such suffixes, all the ones collected in 3.1 are deleted from those collected in 3.2.1. Table 4 shows an example of the most frequent words left after the deletion.

[1] The *Classified Lexical Table Enlarged Edition Monitoring Version* is a thesaurus consisted of 70,858 words under 10,334 sub classes under 842 classes. Accordingly, each sub class consists of approximately 7 words on average.

Table 2: List of collected personal suffixes (top 19).

		Word	Frequency
1	者	person	75
2	氏	Mr.	57
3	さん	Mr., Ms., ...	51
4	会長	president	32
5	長	chief	26
6	議員	assembly person	21
7	人	person	16
7	教授	professor	16
9	家	-er (as in Photographer)	15
10	幹部	executive	14
11	首相	prime minister	13
12	被告	defendant	10
13	業者	trader	9
14	長官	director	8
14	相	minister	8
14	書記長	secretary general	8
14	客	guest	8
14	員	member	8
19	代表	representative	7

Table 3: Numbers of collected personal suffixes: listed and unlisted words are those listed and unlisted in the *Classified Lexical Table*. Classes and sub classes indicate those of which the listed words belong to.

	Originally Collected (3.1.1)	After Expansion (3.1.2)
Listed Words	117	1,776
(Sub Classes)	125	125
(Classes)	59	59
Unlisted Words	19	19
Total Vocabulary	136	1,795

Table 4: List of collected locational / organizational suffixes (Top 20 after deleting personal suffixes).

		Word	Frequency
1	市	city	163
2	内	inside	108
3	側	side	78
4	大会	convention	60
5	県	prefecture	46
6	市内	inside a city	43
7	戦	battle	39
7	政府	government	39
7	駅	station	39
7	会談	meeting	39
11	問題	problem	38
12	署	station, office	36
13	会議	conference	35
14	地方	region	31
15	区	ward	30
16	大震災	severe earthquake	29
16	国内	inside a country	29
18	選	election	28
19	地裁	regional court	27
20	町	town	25

Table 5: Numbers of collected locational / organizational suffixes: listed and unlisted words are those listed and unlisted in the *Classified Lexical Table*. Classes and sub classes indicate those of which the listed words belong to.

	Originally Collected (3.2.1)	After Deletion (3.2.2)	After Expansion (3.2.3)
Listed Words	674	607	7,819
(Sub Classes)	764	697	697
(Classes)	318	307	307
Unlisted Words	91	89	89
Total Vocabulary	765	696	7,908

Table 6: List of collected temporal suffixes (top 19).

		Word	Frequency
1	ため	for the sake of	465
2	時代	era	145
3	中	during, inside	128
4	間	while, between	123
5	前	before, in front of	117
6	時	when, at, hour	116
7	ところ	place, at the point of	79
8	うち	while, inside	65
9	際	when	54
10	場合	in case of, when	47
11	上	after, on	46
12	直後	immediately after	45
13	以来	since then	43
14	日	day, date	40
15	後	after	31
16	ほか	else	30
17	ころ	time, around	27
18	戦後	postwar	25
19	当時	then, in those days	24
19	時期	time, period	24
19	まま	occasionally	24

Table 7: Numbers of collected temporal suffixes: listed and unlisted words are those listed and unlisted in the *Classified Lexical Table*. Classes and sub classes indicate those of which the listed words belong to.

	Originally Collected (3.3.1)	After Expansion (3.3.2)
Listed Words	136	2,124
(Sub Classes)	213	213
(Classes)	104	104
Unlisted Words	20	20
Total Vocabulary	156	2,144

3.2.3 Expansion of Vocabulary

Although 696 suffixes were left after the deletion, they are still insufficient to apply to general caption analysis. Thus we expanded the vocabulary by referring to the *Classified Lexical Table* as described in 3.1.2.

Table 5 shows the number of words collected by this process. In the case of words listed in the *Classified Lexical Table*, the numbers of sub classes and classes, of which the words belong to are also shown.

3.3 Collecting Temporal Suffixes

Here, rules and results of collecting *temporal suffixes*; suffixes that indicate time, such as '— after', and individual nouns such as 'today', are discussed.

3.3.1 Collection from Corpora

Temporal suffixes were collected according to the following rules: search for the following pattern, and extract the underlined word.

`Temporal Noun` or `Adverbial Noun`
Certain `Post-positional Particles`

Following is an example of the collection following this rule. The underlined word is collected as a temporal suffix:

春	(spring)	`Temporal Noun`
から	(from)	`Post-positional Particle`

Table 6 shows an example of the most frequent words collected after the deletion.

3.3.2 Expansion of Vocabulary

Although 156 suffixes were collected, they are still insufficient to apply to general caption analysis. Thus we expanded the vocabulary by referring to the *Classified Lexical Table* as described in 3.1.2.

Table 7 shows the number of words collected by this process. In the case of words listed in the *Classified Lexical Table*, the numbers of sub classes and classes, of which the words belong to are also shown.

4 Evaluation of the Dictionaries Through Television News Caption Analysis

As an evaluation of the collected suffixes (dictionaries), we applied them to actual television news caption analysis. The captions used for the evaluation were those previously mentioned; 2,549 captions that appeared in approximately 370 minutes of news video. Title captions and those that appear in explanative flow-chart figures were omitted, since they are usually not noun phrases and that they could easily be distinguished by image processing. Note that correct answers were provided by a third person.

4.1 Analysis Procedure

Caption analysis was performed as shown in Figure 1. Captions were manually written down for this experiment. Japanese morphological analysis system JUMAN [JUMAN] was employed for morphological analysis. As JUMAN returns the boundaries of morphemes and/or parts of speech, the last morpheme of a caption phrase or sentence is compared with the words in each dictionary. Since JUMAN itself has a proper noun dictionary and a temporal noun dictionary, proper nouns and some temporal nouns were analyzed by JUMAN.

Table 8: Result of personal caption analysis: N_c, N_m, and N_o stand for numbers of correct, mistaken, and oversight answers, f_p and f_r stand for precision and recall, respectively.

	N_c	N_m	N_o	f_p	f_r
Proposed method	260	63	—	80.50%	69.33%
JUMAN	48	54	—	47.06%	12.80%
Combined	308	117	67	72.47%	82.13%

Table 9: Result of locational / organizational caption analysis.

	N_c	N_m	N_o	f_p	f_r
Proposed method	428	571	—	42.84%	53.50%
JUMAN	279	13	—	95.55%	34.88%
Combined	707	584	93	54.76%	88.38%

Table 10: Result of temporal caption analysis.

	N_c	N_m	N_o	f_p	f_r
Proposed method	165	259	—	38.92%	82.50%
JUMAN	22	0	—	100.00%	11.00%
Combined	187	259	13	41.93%	93.50%

4.2 Analysis Result

The result of the analysis is shown in Tables 8 through 10. The numbers listed in 'Proposed method' indicate common nouns analyzed by the proposed method, and those listed in 'JUMAN' indicate proper nouns analyzed by JUMAN's proper noun dictionary, except for temporal nouns. For temporal nouns, the numbers listed in 'JUMAN' indicate those analyzed in JUMAN's temporal noun dictionary. Finally, 'Combined' indicates the overall analysis ability of the combined analysis of both common and proper noun captions. Precision and recall are defined as follows:

$$Precision(f_p) = \frac{Correct(N_c)}{Correct(N_c) + Mistaken(N_m)}$$

$$Recall(f_r) = \frac{Correct(N_c)}{Correct(N_c) + Oversight(N_o)}$$

Precision of each method stands for individual precision rates, and recall of each method stands for the contribution to the overall recall rate.

4.3 Examination

The results showed high accuracy in recall, which is important when finding index candidates with specific semantic attributes from restrictions from other processes, such as image analysis. As a whole, the precision of the proposed method alone is not sufficient.

Nonetheless, as for personal caption analysis, the proposed method and the overall result showed high accuracy both in precision and recall. As for Temporal caption analysis, not for the proposed method, the overall recall would have been extremely low.

There were five major reasons for mistakes and oversights. Table 11 shows the reasons and their ratios. Reasons (a-1) and (d) directly reflects the quality of the collected dictionary, although about two fifth of the mistakes could have been prevented if the compound noun phrase

Table 11: Reasons for mistakes and oversights.

	Reason	Ratio
(a-1)	Lack of common noun vocabulary in the collected dictionary.	44.0%
(a-2)	Lack of proper noun / temporal noun vocabulary in JUMAN's dictionary.	38.8%
(b)	Semantic diversity.	9.1%
(c)	Mistake in morphological analysis by JUMAN.	5.7%
(d)	Noise in the collected dictionary.	2.4%

were divided into granular morphemes. This could be solved by performing exact tail matching between the compound noun phrase and the words in the dictionary, but may cause an increase in the number of mistakes. Reason (c) is an essential problem when dealing with semantic analysis. It is almost impossible to analyze them individually, without other information. As for reasons (a-2) and (b), we will have to wait for JUMAN's improvement.

We will try to refine the dictionary by tackling on reasons (a-1) and (d) in the future.

5 Conclusion

We have introduced a semantic analysis method to television news captions *i.e.* compound noun phrases. The method showed high recall in an evaluation by actual captions derived from television news programs. Although precision of the proposed method is not sufficiently high, we have analyzed the reasons for mistakes and oversights for future refinement of the dictionary.

We will also integrate the analysis method to the news video indexing mechanism to see the efficacy in terms of actual indexing.

Acknowledgments

The Classified Lexical Table (Enlarged Edition Monitoring Version) was used for monitoring under a licensed agreement with National Language Institute of Japan. RWC text database is a product of Real World Computing Partnership (RWCP), and was used under a licensed agreement.

We thank Hiroshi Taira and Tomoe Oda for their laborious work on caption transcription, Masaaki Honda and Ryoji Kobayashi for their help to the evaluation experiment.

References

[CLT] National Language Research Institute of Japan; "The Classified Lexical Table (Enlarged Edition) [Monitoring Version]"; Mar 1996.

[Hisamitsu97] Hisamitsu, T. and Niwa, Y.; "Acquisition of Personal Names from Newspaper Articles by Lexical Knowledge and Co-occurrence Analysis"; *Tech. Rep. IPS Japan* 97-NL-118, pp.1-6, Mar 1997 (*in Japanese*).

[Hori99] Hori, O.; "A Video Text Extraction Method for Character Recognition"; *Proc. Fifth Intl. Conf. on Document Analysis and Recognition (ICDAR'99)*; Sep 1999 (to appear).

[Ide98] Ide, I., Tanaka, H.; "Automatic Semantic Analysis of Television News Captions"; *Proc. Third Intl. Workshop on Information Retrieval with Asian Languages*, pp.56-60, Oct 1998.

[Ide99] Ide, I., Yamamoto, K., and Tanaka, H.; "Automatic Video Indexing Based on Shot Classification"; *Advanced Multimedia Content Processing -First Intl. Conf. AMCP'98, Osaka, Japan-*, S. Nishio, F. Kishino eds., Lecture Notes in Computer Science vol.1554, Springer-Verlag, Mar 1999.

[JUMAN] Language Media Lab., Grad. School of Informatics, Kyoto Univ.; "Japanese Morphological Analysis System JUMAN version 3.6"; Nov 1998.
Available from http://www-lab25.kuee.kyoto-u.ac.jp/nl-resource/juman-e.html

[KUC] Language Media Lab., Grad. School of Informatics, Kyoto Univ.; "Kyoto University Corpus version 2.0"; Jun 1998.
Available from http://www-lab25.kuee.kyoto-u.ac.jp/nl-resource/corpus.html

[Nakamura97] Y. Nakamura, and T. Kanade, "Semantic Analysis for Video Contents Extraction –Spotting by Association in News Video–"; *Proc. Fifth ACM Intl. Multimedia Conf.*, pp.393-402, Nov 1997.

[Nasukawa97] Nasukawa, T.; "Keyword Categorization Based on Discourse Information"; *Proc. Eleventh Annual Conf. JSAI*, pp.348-349, Jun 1997 (*in Japanese*).

[NEtask] United States Defense Advanced Research Projects Agency (DARPA), Information Technology Office; "Named Entity Task Definition, Version 2.1"; *Proc. Sixth Message Understanding Conference*, pp.317-332, Nov 1995.

[RWC] Real World Computing Partnership (RWCP); "RWC Text Database"; Mar 1996.

[Satoh99] S. Satoh, Y. Nakamura, and T. Kanade; "Name-It: Naming and Detecting Faces in News Videos"; *IEEE MultiMedia*, vol.6, no.1, pp.22-35, Mar 1999.

[Sawaki98] M. Sawaki, and N. Hagita; "Text-line Extraction and Character Recognition of Document Headlines with Graphical Designs Using Complementary Similarity Measure", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.20, no.527, pp.91-98, Jan 1998.

[Smith97] Smith, M. A. and Kanade, T.; "Video Skimming and Characterization Through the Combination of Image and Language Understanding Techniques"; *CMU Tech. Rep.* CMU-CS-97-111, Feb 1997.

[Wactlar96] Wactlar, H. D., Kanade, T., Smith, M. A. and Stevens, S. M.; "Intelligent Access to Digital Video: Informedia Project"; *IEEE Computer*, vol.29, no.3, pp.46-52, May 1996.

[Wactler98] Wactler, H. D., Hauptmann, A. G. and Witbrock, M. J.; "Informedia News-on-Demand: Using Speech Recognition to Create a Digital Video Library"; *CMU Tech. Rep.*, CMU-CS-98-109, Mar 1998.

[Watanabe96] Watanabe, Y., Okada, Y. and Nagao, M.; "Semantic Analysis of Telops in TV Newscasts"; *Tech. Rep. IPS Japan* 96-NL-116, vol.96, no.89, pp.107-114, Nov 1996 (*in Japanese*).