TERM KNOWLEDGE ACQUISITION USING THE STRUCTURE OF HEADLINE SENTENCESFROM INFORMATION EQUIPMENTS OPERATING MANUALS

Makoto Imamura, Yasuhiro Takayama

Information Technology R&D Center, Mitsubishi Electric Corporation
5-1-1 Ofuna, Kamakura, Kanagawa, 247-8501, Japan

Masanori Akiyoshi, Norihisa Komoda

Graduate School of Information Science and Technology, Osaka University
2-1 Yamadaoka, Suita, Osaka, 565-0871, Japan

Abstract: This paper proposes a method for automatically extracting term knowledge such as case relations and IS-A relations between words in the headline sentences of operating manuals for information equipments. The proposed method acquires term knowledge by the following iterative processing: the case relation extraction using correspondence relations between surface cases and deep cases; the case and IS-A relation extraction using compound word structures; the IS-A relation extraction using correspondence between the case structures in the hierarchical headline sentences. The distinctive feature of our method is to extract new case relations and IS-A relations by comparison and matching the case relations extracting from the super and sub headline sentences using the headline hierarchy. We have confirmed that the proposed method has achieved 92.4% recall and 96.8% precision for extracting case relations, and 93.9% recall and 89.9% precision for extracting IS-A relations from an operating manual of a car navigation system.

1 INTRODUCTION

As the functions of information equipments increase diversity and complexity in recent years, users can hardly find the target items from several hundred pages of operating manuals. Especially for information equipments such as car navigation systems, mobile phones, and DVD recorders, the help of operation is important because the target users of those are broad range of people (Waida, 2007). To facilitate the help of equipment operations and manual search, the function of refining user's intention is desired, and term knowledge is required to implement the function (Kurohashi, 2000), (Kawahara, 2003). The representative term knowledge are case relations like equipment operations and their objects, and the super-sub relations (IS-A relations) between objects of the equipments. For example, by using term knowledge of case relations and IS-A relations, help systems can suggest the related terms with the user-input keywords to improve search ability. However, the

human workload of term knowledge creation with expertism for each special domain is at issue. Therefore, the needs of term knowledge acquisition from operating manuals, necessarily with the commercialization of the equipments, are growing.

Conventional technologies of acquisition of case relations and IS-A relations are based on statistical machine learning (Oishi, 1997), (Kawahara, 2000) or sentence pattern rules (Kurohashi, 1992), (Kobayashi, 2004). The former methods require huge corpus. The targets of latter methods are restricted in dictionaries and blogs. Therefore, both methods cannot be applied to term knowledge acquisition from operating manuals whose volumes are relatively small. Moreover, the conventional methods are based on one sentence processing, so the methods using the document structure such as the hierarchy of headlines or the relations between the cells in the tables, specific to the business documents, has not been examined.

We focus on the headline hierarchy as the document structure in the business documents. This

paper proposes a method of automatically acquiring case relations and IS-A relations, whose distinctive feature is the matching the case elements by using the headline hierarchy.

2 TERM KNOWLEDGE OF ACQUISITION

In the operation help and the manual search for information equipments, the system can support the users of the equipments to find the desired explanation pages quickly by suggesting the related terms of the user-input keywords using term knowledge. For example, the user can easily reach the targeted page by only choosing the desired terms from the alternatives suggested by the system using term knowledge shown in Table 1, instead of wandering around looking for a number of the pages in the search result list.

Table 1: Examples of term knowledge.

Ex. 1	{verb/settei-suru (setting), object/*mokuteki-chi* (destination place), instrument/*navi-meyu* (navigation menu)} {verb/settei-suru (setting), object/*mokuteki-chi* (destination place), instrument/ *touroku-chi* (registered place)} {verb/*kensaku-suru* (search), object/ *mokuteki-chi* (destination place), instrument/*jusho* (address)}
Ex. 2	IS-A(*kisetsu-no-supotto* (seasonal spot), *osusume-jouhou* (recommended information)) IS-A(*hyakusen*(the hundred best), *osusume-johou* (recommended information)) IS-A(*osusume-supotto* (recommended spot), *osusume-jouhou* (recommended information))

Ex.1 in Table 1 are called the case frames that represent the verb centred relations about the role (deep case) of the noun to the verb. For example, {verb/*settei-suru* (setting), object/*mokuteki-chi* (destination place), instrument/*navi-menyu* (navigation menu)} represents "noun '*mokuteki-chi* (destination place)' has a role of 'object (objective case)', and noun '*navi-menyu* (navigation menu)' has a role of 'instrument (instrumental case)' for the verb '*settei-suru* (setting).'" The case particles with nouns that depend on the verbs are called 'surface cases'. For example, "*tourokuchi-wo* (registered place)" is called "*wo*-case", "*risuto-ni* (to the list)" is called "*ni*-case". From Ex.1 in Table 1, we can see the following: "The verbs with '*mokuteki-chi*

(destination place)' as the object are '*settei-suru* (setting)' and '*kensaku-suru* (search)'"; "The instruments for '*mokuteki-chi-wo settei-suru* (setting the destination place)' are '*navi-menyu* (navigatioin menu)' and '*touroku-chi* (registered place).'"

Ex.2 in Table 1 are examples of IS-A relations extracted from the super-sub relations between terms. These examples represent: "there are 3 instances for '*osusume-jouhou* (recommended information)': '*kisetsu-no supotto* (seasonal spot)', '*hyaku-sen* (the hundred best)', and '*osusume-supotto* (recommended spot).'"

3 RELATED WORK

The methods for automatic extraction of term knowledge are mainly the named entity extraction method (Sudo 2001) and the corpus-based knowledge extraction method (Utsuro 1998). These methods are classified broadly into the machine learning based methods and the linguistic structure patterns based methods.

The machine learning based methods require the huge workload for preparation, because they require huge corpus like hundreds of millions of sentences or the training data created by human. Then, there are the problems: (1) It is difficult to collect huge corpus for certain business domains. (2) The cost for creating answer data is high. Therefore, it is hard to apply the methods to acquire technical term knowledge in business domains.

The linguistic structure patterns based methods enable term knowledge acquisition to decrease the required size of documents by using the human-created pattern rules specified in each type of documents. The examples of these methods are "term knowledge extraction of hyper text construction from dictionary (Kurohashi 1992)", and "term knowledge acquisition for sentiment extraction (Kobayashi 2004)." The former specialize the target documents for extraction in the dictionary and the latter specialize those in Web bulletin boards and blogs. Because the pattern rules need to be created for each type of documents, the specific pattern rules are difficult to apply to the other type of documents. We call this issue "issue 1: using characteristics of the headline sentences in the information equipment manuals."

Moreover, the conventional methods are based on processing one sentence, so the case relations and IS-A relations acquisition using the headline hierarchy has not been examined. We call this issue "issue 2: case relations and IS-A relations in the

headline hierarchy." The conventional methods use the surface dependency relations with case particles, so extracting the case relations between words that consist of the compound words has not been examined. We call this issue "issue 3: case relations in the compound words."

4 TERM KNOWLEDGE ACQUISITION USING THE HEADLINE STRUCTURE

4.1 Overview of Proposed Method

To cope with "issue 1: using characteristics of the headline sentences in the information equipment manuals", we propose a method of extracting case relation by the correspondence rules between surface and deep case based on the speciality of linguistic expressions of operations, status and conditions in the operating manuals. We also propose a method of extracting IS-A relation by using the commonality of word ending in compound words.

To cope with "issue 2: case relations and IS-A relations in the headline hierarchy", we propose a method of extracting case relation by merging cases in the super and sub headlines when the both verbs are identical, but cases are different. We also proposes a method of extracting IS-A relations, by comparing with cases in the super and sub headlines when the both verbs are the identical, and both cases are also identical.

To cope with "issue 3: case relations in the compound words", we propose a case analysis method for compound words by using the linguistic expressions specific in the operating manuals.
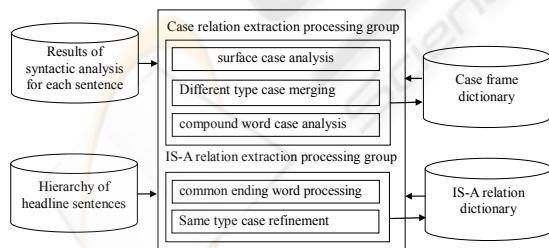

Figure 1: The architecture of the automatic term knowledge acquisition method.

The architecture of our proposed term knowledge acquisition consists of the processings corresponding to the above methods, the case frame dictionary that is a set of case relations, and the IS-A relation dictionary that is a set of IS-A relations as shown in Figure 1. Each processing extracts new case

relations or IS-A relations by using the results of syntactic analysis for each sentence, the hierarchy of headline sentences, the case frame dictionary, and the IS-A relation dictionary, then registers the new relation in the case frame dictionary or the IS-A relation dictionary. The whole process of the proposed method is repeated iteratively until no new term knowledge can be extracted.

4.2 Extraction of Case Relations

4.2.1 Case Relations

There is no consensus on what deep cases exist (Tsujii 1985), therefore, we use the deep cases shown in Table 2, which are customized from Fillmore's deep cases (Fillmore 1975) for the operations of information equipments.

Table 2: Deep cases for information equipment operations.

Deep case	Explanation
Agentive case (A)	Information equipment of operation target, or the parts of it
Experiencer case (E)	User of the equipment
Objective case (O)	Operation target of the users
Instrumental case (I)	Instrument or mean of equipment operation
Source case (S)	Status or place before equipment operation
Goal case (G)	Status or place after equipment operation
Locative case (L)	Place or location about equipment operation
Time case (T)	Condition of equipment operation

Table 3: Examples of case relations in the headlines.

Headline sentence	Case relation
touroku-chi-no sentaku (choice of the registered place)	{verb/*sentaku-suru* (choose), object/*touroku-chi* (registered place)}
koe-niyoru sousa (operation by voice)	{verb/*sousa-suru* (operate), instrument/*koe* (voice)}
Saisei-gamen-kara navi-gamen-ni modosu (bring replay screen back to the navigation screen)	{verb/*modosu* (bring back), source/*saisei-gamen* (replay screen), goal/*navi-gamen* (navigation screen)}

Table 3 shows the examples of case relations obtained by analysing the headline sentences in the operating manuals of a car navigation system. Here we denote the cases by attribute-value pairs whose attribute names are the deep cases and values are the words or phrase in the sentences. For example, {verb /*sentaku-suru* (choose), object/*touroku-chi*

(registered place)} represents "the objective case of the verb *sentaku-suru* (choose) is *touroku-chi* (registered place)."

4.2.2 Surface Case Analysis

The corresponding relations between surface and deep cases become simple when we restrict the sentences to the headlines of operating manuals for equipments. The processing for the correspondence between surface and deep case extracts the case relations by determining deep cases from the analyzed results of each sentence using the rules shown in the left and the upper-right column in Table 4. The underlined expressions in Table 4 are examples of surface cases.

In Table 4, the operation words, status words, and conditional expressions, specific to the operating manuals of information equipments, are as follows.

(i) Operation Words. The operation words are the verbs and *sahen*-nouns in the headline sentences. For example, "*settei* (setting)" in "*mokuteki-chi-no settei* (setting of the destination place)" and "*meiru* (mail)" in "*meiru kinou* (mail function)" are the verbs, and then they are the operation words.

(ii) Status Words. The status words are nouns that are related by the verbs with the expressions of beginning, duration, and completion. For example, "*sukurouru-suru* (scroll)" is an operating word, auxiliary verb "*ta* (have -ed)" represents completion of (9) in Table 4: "*sukurouru-shita chiten* (the place that have been scrolled to)", so the noun "*chiten* (the place)" is a status word. The examples of beginning, duration, and completion are "*hajimeru* (start)", "*teshimau* (have -ed)", and "*teiru* (have been)". We can make the dictionary beforehand for the status words, because these expressions are formulaic and domain independent.

(iii) Conditional Expressions. The conditional expressions are terms that represents conditions such as "*baai* (case)", "*ji* (time)", and "*sai* (as)". The conditional expressions have functions as the surface cases. For example, the "*ji* (time)" in "*setsuzoku-ji-no kakunin* (confirmation at the time of connection)" of (10) in Table 4 is a conditional expression, and "*ji-no* (at the time of connection)" has a function of time case. We can make the dictionary beforehand for the condition words, because these expressions are formulaic and domain independent.

Table 4: The correspondence rules between surface and deep case.

Deep case	Surface case that can be the left deep case
	Example of headline sentence
(O)	- *wo*-case
	- *no*-case that is *sahen*-noun or adverbial form of verb that cannot be time case
	(1) *shisetsu-wo sagasu* (search the facility)
	(2) *touroku-chi-no sentaku* (choice of the registered place)
(I)	- *niyoru*-case
	- *deno*-case or *no*-case that cannot be (T) or (L)
	- *kara*-case that cannot be source case
	(3) *koe-niyoru sousa* (operation by voice)
	(4) *jusho-de sagasu* (search by the address)
	(5) *risuto-kara sagasu* (search by the list)
(S)	- *kara*-case or *karano*-case whose depending operation word is depended by (G)
	(6) *saisei-gamen-kara navi-gamen-ni modosu* (bring the replay screen back to the navigation screen)
(G)	- *ni*-case or *heno*-case
	(7) *genzai-chi-ni modosu* (bring back to the current place)
	(8) *audio-kinou-heno setsuzoku* (connection to the audio function)
(L)	- *deno*-case or *de*-case with a status word
	(9) *sukurouru-shita chiten-deno sousa* (operation at the place that have been scrolled)
(T)	- *deno*-case or *no*-case following a conditional expression
	(10) *setsuzoku-ji-no kakunin* (confirmation at the time of connection)

(O): objective case, (I): instrumental case, (S): source case , (G): goal case, (L): locative case, (T): time case

4.2.3 Different Type Case Merging

Different type case merging extracts case relations by merging the different cases in the case relations between the super and sub headlines when the both verbs are the same, using rules shown in Table 5.

Table 5: The merge rules for different type cases.

Category	The merge rules for different type cases
With case element	if (ec(Si) = {verb/X, C1/A} $\wedge$ ec(Sj) = {verb/X, C2/B} $\wedge$ headline-hiearchy(Si, Sj)) then ec(Sj) = { verb/X, C1/A, C2/B }
Case element ellipsis	if (ec(Si) = {verb/X, C1/A} $\wedge$ ec(Sj) = { verb/Y } $\wedge$ headline-hiearchy(Si, Sj)) then ec(文 j) = { verb/Y, C1/A }

For example, this processing can extract the case relations "{verb/*settei-suru* (setting), object/*touroku-chi* (registered place), instrument/*kensaku-rireki* (search history)}" and "{verb/*settei-suru* (setting), object/*touroku-chi* (registered place), instrument/*jusho* (address)}" form headline hierarchy of the left column in Table 6.

Table 6: Case relations extracted from each headline.

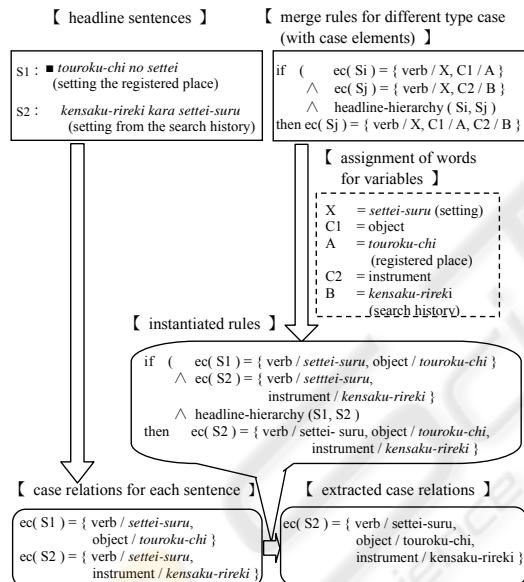| Headline sentence	Case relation
■ *touroku-chi- no settei* (setting the registered place)	{verb/*settei-suru* (setting), object/*touroku-chi* (registered place)}
- *kensaku-rireki- kara settei-suru* (setting from search history)	{verb/*settei-suru* (setting), instrument/*kensaku-rireki* (search history)}
- *jusho-kara settei-suru* (setting from the address)	{verb/*settei-suru* (setting), instrument/*jusho* (address)}

Figure 2: Extraction of case relations by the merge rules for different type cases.

The execution flow of the different type case merge processing is shown in Figure. 2. Here, S1 denotes the headline in the upper layer (super headline), S2 denotes the headline in the lower layer (sub headline). This processing merges the deep case "C1/A = object/*touroku-chi* (registered place)" in S1 with the deep case "C2/B = instrument/*kensaku-rireki* (search history)" in S2, using the common verb "X = settei-suru (setting)" in the case relations extracted from each sentence. Then the processing extracts the case relation {verb/*settei-suru* (setting), object/*touroku-chi* (destination place),

instrument/*kensaku-rireki* (search history)} by complement with the objective case omitted in S2.

4.2.4 Compound Word Case Analysis

Compound word case analysis extracts case relations form compound words. By using the rules in Figure 3, a new case relation {verb/Mn, case C/[M1, M2, …, Mn-1]} is extracted, if the word ending is an operation word Mn, and the noun [M1, M2, …, Mn-1] is in the case frame dictionary as a case element or a headline word.
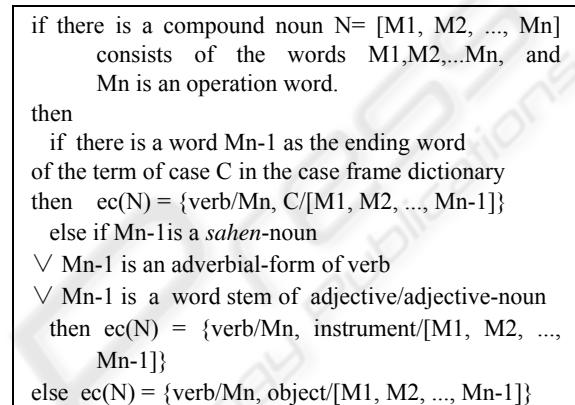
```
if there is a compound noun N= [M1, M2, ..., Mn]
        consists of the words M1,M2,...Mn, and
        Mn is an operation word.
then
   if there is a word Mn-1 as the ending word
of the term of case C in the case frame dictionary
then   ec(N) = {verb/Mn, C/[M1, M2, ..., Mn-1]}
   else if Mn-1is a sahen-noun
∨ Mn-1 is an adverbial-form of verb
∨ Mn-1 is a  word stem of  adjective/adjective-noun
   then   ec(N) = {verb/Mn, instrument/[M1, M2, ...,
       Mn-1]}
else  ec(N) = {verb/Mn, object/[M1, M2, ..., Mn-1]}
```

Figure 3: The case analysis rules for compound words.

Table 7: Examples of case relations extracted from compound words in headlines.

	Headline sentence
	Extracted case relation
Ex. 1	(1) *sisutemu-onryo- no henkou* (modification of volume in the system) (2) *AV-onryo settei* (AV-volume setting)
	{verb/*henkou-suru* (modify), object/[*sisutemu* (system), *onryou* (volume)]} {verb/*settei-suru* (setting), object/[*AV* (AV), *onryou* (volume)]}
Ex. 2	(1) *ukai-ruuto-wo tsukuru*(create a detour route) (2) *ruuto-henkou* (change the route)
	{verb/*tsukuru* (create), object/[*ukai* (detour), *ruuto* (route)] } {verb/*henkou* (change), object/[*rouuto* (route)]}

Table 7 shows examples of case relations extracted from compound words. If the headline sentence "*AV-onryo settei* (AV-volume setting)" is a compound noun N, N= [*AV* (AV), *onryo* (volume), *settei* (setting)], and the operating word is Mn = "*settei* (setting)". Then, Mn-1= "*onryo* (volume)" is the ending word of the case "C = object" in the case relation extracted from the headline sentence "*sisutemu-onryo-no henkou* (modification of volume in the system)." Therefore, a new case relation

{verb/*settei-suru* (setting), object/[*AV* (AV), *onryou* (volume)]} can be extracted by the rules in Figure 3.

4.3 Extraction of IS-A Relations

4.3.1 Common Ending Word Processing

There is a conventional heuristics that an IS-A relation holds when an inclusive relation as substring between words that consists of compound words. This heuristics incorrectly extracts IS-A(*settei-naiyou* (setting content), *naiyou* (content))" whose IS-A relation is not important for applications like the operation help. Common ending word processing extracts IS-A relations between terms by the rule using the headline suffixes and the case frame dictionary that are the specific expressions for the headlines in the manuals.

Here, the headline words are the headline sentences whose endings are the nouns depending on "the headline suffixes" such as "*gaiyou* (overview)", "*toha* (what ~is)", "*oteire* (maintenance)", and "*sousa* (operation)", or the nouns without the headline suffixes. For example, "*onryou-settei-no gaiyou* (overview of volume setting)" is a headline sentence, "*onryou-settei* (volume setting)" is a headline word because "*gaiyou* (overview)" is a headline suffix. We can make the dictionary of the headline suffixes beforehand, because they are the formulaic and domain independent.

```
if (there is a compound noun N= [M1, M2, ..., Mn]
consists of the words M1,M2,...Mn
∧ there is a word Mn-1
        as the ending word of the term of case C
        in the case frame dictionary,
        or the ending word of the headline word )
  then IS-A(N, Mn)
```

Figure 4: The rule for the words with the common ending.

Using the rule in Figure 4, "IS-A(*AV-onryo* (AV-volume), *onryo* (volume))" can be extracted from "*Sisutemu-onryo-wo henkou-suru* (modify the system volume)" and "*AV-onryo* (AV-volume)." Because "{verb/ *henkou-suru* (modify), object/[*sisutemu* (system), *onryou* (volume)]} " is extracted from *sisutemu-onryo-wo henkou-suru* (modify the system volume)", here, "compound noun N=[*AV* (AV), *onryo* (volume)]", "Mn= *onryo* (volume)", and "C=object", then "IS-A(*AV-onryo* (AV-volume), *onryo* (volume))" is extracted.

4.3.2 Same Type Case Refinement

Same type case refinement extracts the IS-A relations using the case refinement rules for the same type cases for nouns and for verbs in Table 8.

Table 8: The case refinement rules for the same type cases.

Category	The case refinement rules for the same type cases
For nouns	if (ec(Si) = {verb/X1, C1/A} ∧ ec(Sj) = {verb/X2, C2/B} ∧ (X1 = X2 ∨ IS-A(X2, X1)) ∧ headline-hiearchy(Si, Sj)) then IS-A(B, A)
For verbs	if (ec(Si) = {verb/X1, C1/A1} ∧ ec(Sj) = {verb/X2, C2/A2} ∧ (A1 = A2 ∨ IS-A(A2, A1)) ∧ headline-hiearchy(Si, Sj)) then IS-A(Y, X)

For example, by using the case refinement rules for the same type cases for nouns, IS-A(*kisetsu-no spotto* (seasonal spot), *osusume-jouhou* (recommended information)) and IS-A(*hyaku-sen* (the hundred best), *osusume-jouhou* (recommended information)) are extracted from headline sentences "'*kisetsu-no supotto-wo sagasu* (search the seasonal spot) 'and '*hyakusen-wo sagasu* (search the hundred best)'" which have super headline '*osusume-jouhou-wo sagasu* (search the recommended information).'"

5 EVALUATION

5.1 Experimental Setting

We describe the evaluation result of our proposed method using "an instruction manual for a car navigation system (Mitsubishi 2005)" as input data. The input data are 875 headline sentences with hierarchical tags, and the output data are case relations and IS-A relations. Here we use JUMAN (Japanese Morphological Analysis System) (Kurohashi, 2005a) and KNP (Kurohashi-Nagao Parser) (Kurohashi, 2005b) for sentence analysis. In evaluation, we calculate recall and precision by comparing with the human-extracted relations as answer relations. The numbers of answer relations are 858 for case relations and 495 for IS-A relations.

306

5.2 Recall and Precision

We have compared that the proposed method using headline hierarchy relation with the basic method using one sentence analysis. Here we consider the processing for the correspondence between surface and deep case as the basic method for case extraction, and the processing for the words with the common ending as the basic method for IS-A relation, because they are considered to be a customization of the conventional methods to the operating manual headlines. The evaluation result for extracting case relations of the proposed method has achieved 92.4% recall and 96.8% precision, increasing 31.1% and 1.5% respectively from the basic method as shown in Table 9.

The evaluation result for extracting IS-A relations has achieved 93.9% recall and 89.9% precision, increasing 21.0% and 1.0% respectively from the basic method as shown in Table 10.

Table 9: An evaluation result of case relation extraction.

	Basic	Proposed	Improvement
Recall	61.3%	92.4%	31.1%
Precision	95.3%	96.8%	1.5%

Table 10: An evaluation result of IS-A relation extraction.

	Basic	Proposed	Improvement
Recall	72.9%	93.9%	21.0%
Precision	88.9%	89.9%	1.0%

The proposed method has almost achieved 90 % of target recall and precision on the practical view. The future issue is the establishment of the method for suggesting related information like certainty factors for human to correct false extraction.

5.3 Analysis of Extraction Failure

5.3.1 Extraction Leaks for Case Relations

The 22 relations (7.6% of all answers) of 65 total leaks of case relations can be extracted by human from one sentence as shown in Ex.1 and Ex.2 in Table 11. They are difficult to extract because of the lacks of clue in the target sentences. A future issue is to establish a method of finding sentences, which have the clue from the body of the manuals or Web text, such as "*denwa-de tsuuwa-suru* (call by telephone)", "*taju-de moji-wo housou-suru* (broadcast the characters by multiplex)."

The other 43 leak relations can be extracted by human from headline hierarchy. The 30 leak relations in them need knowledge about verbs

regarding the actions and the results of the operations as Ex.3 in Table 12. The 13 leak relations in them need to fill the ellipsis of case elements in the sentences. To cope with former cases, it is required to describe knowledge about basic verbs. To cope with latter cases, it is good to consider the convention of designing the headlines for the operating manuals.

Table 11: Examples of extraction leak for case relations.

	Headline sentence
	Extraction leak for case relation
Ex. 1	*denwa-no tsuwa* (call by telephone)
	{verb/*tsuwa-suru* (call), instrument/*dennwa* telephone)}
Ex. 2	*moji-taju-housou* (teletext broadcast)
	{verb/*housou-suru* (broadcast), object/*moji* (character), instrument/*tajuu* (multiplex)}
Ex. 3	■ *torakku-risuto-wo hyouji-suru* (display the track list) - *arubamu-risuto-kara modoru* (return from the album list)
	{verb/*modoru* (return), source/*torakku-risuto* (track list), goal/*arubamu-risuto* (album list)}
Ex. 4	■ *adoresu-chou-wo kanri-suru* (manage the address book), - *touroku-suru* (register)
	{verb/*touroku-suru* (register), object/*adoresu* (address), goal/*adoresu-chou* (address book)}

5.3.2 Extraction Errors for Case Relations

The extraction errors for case relations are 26 (7.6% of all extracted relations). The 4 errors are caused by the processing for the correspondence between surface and deep case. For example, the case relation {verb/*tsuuwa-suru* (call), object/*denwa* (telephone)} is extracted as an error. The 13 cases are caused by the case analysis processing for the compound word. For example, the case relation {verb/*housou-suru* (broadcast), object/*moji-tajuu* (multiplex)} is extracted as an error of Ex.2 in Table 11. The other 9 errors are caused by the different type case merge processing. For example, the case relation {verb/*tourokur-suru* (register), object/*adoresu-chou* (address book)} is extracted as an error of Ex.4 in Table 11 The countermeasure of them are the same as the extraction leaks.

5.3.3 Extraction Leaks for IS-A Relation

The 30 cases of extraction leaks for IS-A relations (6.1% of all answers) do not meet the preconditions

of the same type case refinement processing as shown in Table 12. It needs using the meaning of the terms to extract them correctly. A future issue is a method of suggesting how to loose the preconditions of the extraction rules.

Table 12: Examples of extraction leak for IS-A relation.

Headline sentence
Extraction leak for IS-A relation
■ *tsuushin-houhou-nitsuite* (on communication method)   - *i-moudo-setsuzoku* (i-mode connection)   - *internettto-setsuzoku* (internet connection)
IS-A(*i-moudo-setsuzoku* (communication method),   *tsuushin-houhou* (i-mode connection))   IS-A(*internetto-setsuzoku*(internet connection),   *tsuushin-houhou* (internet connection))

5.3.4 Extraction Errors for IS-A Relations

The extraction errors for IS-A relations are 52 (10.1% of all extracted relations). The 45 errors of them are caused by the processing for the words with the common ending. For example, the IS-Arelation IS-A(*mokuteki-chi* (destination place), *chi* (place)} is extracted as an error from the headline sentences: "*mokuteki-chi-no kensaku* (search of the destination place)" and "*touroku-chi-wo sagasu* (search the registered place)". This is because the word *chi* (place) that is the ending word of the compound noun *mokuteki-chi* (destination place) is in the case relation {verb/*sagasu* (search), object/ *tourokuchi-chi* (registered place)}. The noun suffixes such as *chi* (place), "*mei* (name)", "*chou* (book)", and "*kata* (type)", that appear in the word ending of the proper nouns or technical terms are unnecessary for extracting terms. It is preventable if we prepare the dictionary for noun suffixes, however, we need to create the noun suffix dictionary depending on each application domain.

6 CONCLUSIONS

This paper describes a method for automatically extracting case relations and IS-A relations from operating manuals for help or search about equipment manuals. The distinctive feature of the proposed method is to use the hierarchical structure of the headline sentences, while the conventional methods are based on the processing of one sentence.

We have confirmed the effectiveness of our method by comparison of extraction accuracy of term knowledge with the conventional methods.

REFERENCES

Waida, R., Hoshino, H., Tanaka, K., Shiratori K., 2007. Navigation type Built-in manual on TV, *Technical Communication Simposium 2007.* (in Japanese).

Kurohashi, S., Higasa, W., 2000. Dialogue Helpsystem based on Flexible Matching of User Query with Natural Language Knowledge Base, *1st SIGdial Workshop on Discourse and Dialogue*, pp. 141-149.

Kawahara T., Ito, R., Komatani, K., 2003. Spoken Dialogue System for Queries on Appliance Manuals using Hierarchical Confirmation Strategy, *Proc. of 8th European Conf. on Speech Communication and Technology (Eurospeech2003)*, pp. 1701-1704.

Oishi, A., Matsumoto, Y., 1997. Detecting the organization of semantic subclasses of Japanese verbs, *International journal of corpus linguistics*, Vol.2, No.1, pp. 65-89.

Kawahara, D., Kanji, N., Kurohashi, S, 2000. Japanese Case Structure Analysis by Unsupervised Construction of a Case Frame Dictionary, *Proc. of the 18th International Conference on Computational Linguistics (COLING2000)*, pp.432-438.

Kurohashi, S., Nagao, M., Sato, S., Murakami, M., 1992. A Method of Automatic Hypertext Construction from an Encyclopedic Dictionary of a Specific Field, *Proc. of Third Conference on Applied Natural Language Processing*, pp. 239-240.

Kobayashi, N., Inui, K., Matsumoto, Y., Tateishi, K., and Fukushima, F., 2004. Collecting Evaluative Expressions for Opinion Extraction, *Proc. of the First International Joint Conference on Natural Language Processing (IJCNLP-04)*, pp. 584-589.

Sudo, K., Sekine, S., Grishman, R., 2001. Automatic Pattern Acquisition for Japanese Information Extraction, *Proc. of the First International Conference on Human Language Technology Research*, pp. 1-7.

Utsuro, T., Matsumoto, Y., 1998. Lexical Knowledge Acquisition from Corpora, *Proc. of JSPS-HITACHI Workshop on New Challenges in Natural Language Processing and its Application*, pp.82-87.

Tsujii, J., Yamanashi, M., 1985. Cases and their accreditation criteria, *IPSJ SIG Notes*, NL52-3, pp.1-8. Information Processing Society of Japan. (in Japanese).

Fillmore, C. J., (translated into Japanese by Tanaka, H., Funaki. M.), 1975. *Toward A Modern Theory Of Case & Other Articles*, Sanseido Bookstore (in Japanese).

Mitsubishi Electric Corporation., 2005. HDD Navigation System CU-H9700 series Instruction Manual. (in Japanese).

Kurohashi, S., Kawahara, D., 2005a. *Japanese Morphological Analysis System JUMAN 5.1.* (in Japanese).

Kurohashi, S., Kawahara, D., 2005b. *KN parser (Kurohashi-Nagao parser) 2.0 Users Manual.* (in Japanese).