Improving Coreference Resolution Using Bridging Reference Resolution and Automatically Acquired Synonyms

Ryohei Sasano*[1], Daisuke Kawahara[2], and Sadao Kurohashi[3]

[1] Graduate School of Information Science and Technology, University of Tokyo,
7-3-1 Hongo Bunkyo-ku, Tokyo, 113-8656, Japan,
ryohei@nlp.kuee.kyoto-u.ac.jp,
[2] National Institute of Information and Communications Technology, Japan
3-5 Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0289, Japan
[3] Graduate School of Infomatics, Kyoto University
Yoshida-honmachi, Sakyo-ku, Kyoto, 606-8501, Japan

Abstract. We present a knowledge-rich approach to Japanese coreference resolution. In Japanese, proper noun coreference and common noun coreference occupy a central position in coreference relations. To improve coreference resolution for such language, wide-coverage knowledge of synonyms is required. We first acquire knowledge of synonyms from large raw corpus and dictionary definition sentences, and resolve coreference relations based on the knowledge. Furthermore, to boost the performance of coreference resolution, we integrate bridging reference resolution system into coreference resolver.

1 Introduction

In text, expressions that refer to the same entity are repeatedly used. Coreference resolution, which recognizes such expressions, is an important technique for natural language processing. This paper focuses on coreference resolution for Japanese text.

In Japanese, pronouns are not used so much; most anaphors are represented as proper noun phrases or common noun phrases. To resolve coreference for such language, string matching technique is useful, because an anaphor and its antecedent often share strings [1]. Learning-based coreference approaches, which have been intensively studied in recent years [2–4], use string matching as features for learning. However, in some cases, coreferential expressions share no string, and string matching technique can not be applied.

Resolving such coreference relations requires knowledge that these two expressions share a same meaning. Then, we first propose a method for extracting

* Research Fellow of the Japan Society for the Promotion of Science (JSPS)

synonyms[4] from large raw corpus and dictionary definition sentences, and utilize the synonyms to coreference resolution.

Our target language Japanese also has a characteristic that it has no article. Articles can be a clue for anaphoricity determination, so this characteristic makes anaphoricity determination difficult. We combine bridging reference resolution with coreference resolution as a clue to determine anaphoricity. Roughly speaking, we consider modified NPs are not anaphoric. But if an NP have a bridging relation, it is considered as anaphoric.

The rest of the paper is organized as follows. In Section 2, we present a method for extracting synonyms from raw corpus and dictionary definition sentences. In Section 3, we present basic strategy for coreference resolution and how to use the extracted synonyms and the result of bridging reference resolution for coreference resolution. We show the experimental results on news paper articles in Section 4 and compare our approaches with some related work in Section 5.

2 Synonym Extraction

It is difficult to recognize coreference relations between absolutely different expressions without knowledge of synonyms. To construct a high-performance coreference resolver, we acquire knowledge of synonyms in advance.

As resources for synonym extraction, we use raw corpus and dictionary definition sentences. The characteristic of synonyms extracted from raw corpus is the ability to respond to new words. However, very familiar synonyms, such as *US* and *America*, is not extracted from parenthesis expressions. Thus, in order to extract very familiar synonyms, we also extract synonyms from dictionaries for humans.

2.1 Synonym Extraction from Parenthesis Expressions

When unfamiliar synonymous expressions are used for the first time in text, the information is often written in text by using parenthesis. In example (1), "*KEDO*", a synonym of "*Chosen Hanto Enerugi Kaihatu Kiko*"(Korean Peninsula Energy Development Organization), is written in the following parenthesis. Therefore, we first extract synonyms from parenthesis expressions that appeared in raw corpus.

(1) *Suzuki Chosen Hanto Energy Kaihatu Kiko* (KEDO)
Suzuki Korean Peninsula Energy Development Organization (KEDO)
taishi-ga yutai-shita.
ambassador retire
(The Korean Peninsula Energy Development Organization (KEDO) Ambassador Suzuki retired.)

[4] In this paper, synonyms include acronyms and abbreviations.

Table 1. Thresholds for synonym extraction.

	type	two-way	threshold
1	One consists of English letters and the other does not	yes	1
		no	50
2	One consists of Japanese letters *katakana* and the other does not	yes	2
		no	300
3	One consists of Chinese characters and the other is the abbreviation [5]	yes	0
		no	infinite
4	others	yes	30
		no	infinite

Parenthesis is not always used to indicate synonym. For example, parenthesis is sometimes used to indicate attribution of preceding noun phrases such as age or affiliation. Thus, the problem is how to extract parenthesis pairs that indicate synonym.

In order to deal with this problem, we make an assumption that if a pair A and B that is appeared in parenthesis expression "*A(B)*" is a synonym pair, the frequency of the parenthesis expressions is high and the reverse pair "*B(A)*" can also appeared in corpus. According to this assumption, we extract synonym pairs from parenthesis expressions as follows:

1. Count the frequency of pairs A and B. B is an expression in a parenthesis and A is the preceding noun phrase, that is "*A (B)*".
2. Set frequency thresholds for several types by observing the frequencies of randomly selected 100 pairs.
3. If the frequency exceeds the thresholds, the pair A and B is judged as a synonym pair.

Table 1 shows the thresholds, which are set not to extract incorrect synonym pairs. When there are also examples of "*A (B)*" besides "*B (A)*", we call this pair as two-way pair, and use the geometric mean of the frequencies against the looser threshold.

We extracted synonym pairs from Japanese newspaper articles in 26 years (12 years of *Mainichi* newspaper and 14 years of *Yomiuri* newspaper). There are about 10 million parenthesis expressions in the newspaper articles.

Table 2 shows the result of extraction. We acquired 2,653 synonym pairs. Almost all of the extracted synonym pairs are correct because we set the threshold not to extract incorrect synonym pairs.

[5] One expression must include all Chinese characters included in the other expression.

Table 2. The result of synonym extraction from parenthesis expressions.

type	#	examples
1	1,572	*kokunai sou-seisan* = GDP domestic gross product GDP
		Europe rengo = EU European Union EU
2	732	*jugyo keikaku* = *syllabus* class plan syllabus
		shien kigyo = *sponsor* support company sponsor
3	239	*Gakushu kenkyuu sha* = *Gakken* study pursuit corporation = Gakken
		Nihon kogyo ginko = *Kogin* Japan industrial bank = Kogin
4	110	*ushi kaimenjou noushou* = *kyogyubyo* bovine spongiform encephalopathy mad cow disease
		Myanmar = *Burma* Myanmar Burma
sum	2,653	

2.2 Synonym Extraction from Dictionary

Secondly, in order to extract very familiar synonyms, we use definition sentences of dictionaries for humans. The following process is carried out for each dictionary entry A.

1. If the definition sentence ends with "*no ryaku*" (abbreviation of) or "*no koto*" (synonym of), we extract the rest of the sentence as a synonym candidate B; otherwise extract whole the sentence as B.
2. If B itself is an entry of dictionaries or enclosed by angle brackets, the pair of A and B is judged as a synonym pair.

We extracted synonyms from *Reikai Shougaku Kokugojiten* [5] and *Iwanami Kokugo Jiten* [6]. As a result, we extracted 402 synonym pairs from dictionary definition sentences. Table 3 shows examples of extracted synonym pairs.

Only 4 synonym pairs extracted from dictionary definition sentences overlapped with the synonym pairs extracted from parenthesis expressions. Therefore, it is reasonable to suppose that we extract very familiar synonyms from definition sentences that were not extracted from parenthesis expressions in raw corpus.

As a whole, we acquired 3,051 synonym pairs from raw corpus and dictionary definition sentences.

3 Strategy for Coreference Resolution

We propose a method to improve coreference resolution using knowledge of synonyms and bridging reference resolution.

Table 3. Examples of extracted synonyms from dictionaries.

type of definition sentence	examples	
	entry	extracted synonym
. . .-no ryaku	*fukei* policewoman	*fujin keikan* woman cop
	Niti JP	*Nihon* Japan
. . .-no koto	*Chuugoku* China	*Chuuka Jinmin Kyowakoku* the People's Republic of China
	Bei US	*America* America
others	*Chokou* Yanzi Jiang	*Yousukou* Chang Jiang
	Japan Japan	*Nihon* Nippon

3.1 Basic Strategy for Coreference Resolution

The outline of our coreference resolver is as follows:

1. Parse input sentences by using a Japanese parser and recognize named entity.
2. Consider each subsequence of a noun phrase as a possible anaphor if it meets "*Condition 1*".
3. For each anaphor:
   (a) From the position of the anaphor to the beginning of document, consider each noun sequence as antecedent candidate.
   (b) If the anaphor and the antecedent candidate meet "*Condition 2*", judge as coreferential expressions and move to next anaphor.

"*Condition 1*" and "*Condition 2*" are varied between methods. "*Condition 1*" judge the anaphoricity of the subsequence.

We use KNP [7] as a Japanese parser. To recognize named entity, we apply a method proposed by Isozaki and Kazawa [8] that use NE recognizer based on Support Vector Machines.

3.2 Determination of Markables

The first step of coreference resolution is to identify the markables. Markables are noun phrases that related to coreference. We consider how to deal with compound nouns.

Previous work on coreference resolution in Japanese focused on the whole compound noun and cannot deal with this example:

(2) *Lifestyle-no chosa-wo jisshi-shita. Chosa naiyo-wa . . .*
lifestyle investigation conduct investigation content
(ϕ conducted an investigation. The content of the investigation was . . .)

In this example, the second "*chosa*"(investigation) that is contained in a compound noun "*chosa naiyo*" refers to the preceding "*chosa*". To deal with such a coreference relation, we consider every subsequence of a compound noun as a markable, that is, we consider "*chosa naiyo*", "*chosa*" and "*naiyo*" as a markable for *chosa naiyo*.

But we consider named entities as an exception. Named entities are not divided and handled as a whole.

3.3 Baseline Methods

We consider 3 baseline methods. In all of these methods, "*Condition 2*" is true when the anaphor exactly matches the antecedent candidate. Only "*Condition 1*" (i.e. anaphoricity determination) varies among these 3 baselines.

In a primitive baseline (*baseline 1*), "*Condition 1*" is always true, that is, every noun sequence is considered an anaphor.

For a bit more sophisticated baselines (*baseline 2* and *baseline 3*), we assume that a modified noun phrase is not anaphoric.

(3) a. *Uno shusho-wa Doitsu-ni totyaku-shita. Shusho-wa kuukou-de ...*
Uno prime minister Germany arrived prime minister airport

(Prime minister Uno arrived in Germany. At the airport the minister ...)

b. *Uno shusho-wa Doitsu-ni totyaku-shita. Asu Doitsu Shusho-tono...*
Uno prime minister Germany arrived Tomorrow German prime minister

(Prime minister Uno arrived in Germany. Tomorrow, with German prime minister ...)

In example (3a), "*shusho*" (prime minister) in the first and second sentence refer to the same entity, but not in example (3b). This is because the second "*shusho*" in (3b) is modified by "*Doitsu*" (German), and this "*shusho*" is turned out to be a person other than "*Uno shusho*".

We consider that a partial noun sequence of a compound noun is modified by its preceding nouns in the compound noun. For example, for the compound noun "XY", "Y" is considered to be modified by "X", and thus "Y" is regarded as non-anaphoric (in this case, noun sequences "XY" and "X" are regarded as anaphoric).

In both *Baseline 2* and *baseline 3*, modified noun phrases are considered non-anaphoric. These two methods differ in the scope of the considered modifier. In *baseline 2*, "*Condition 1*" is true when the noun sequence is not modified by its preceding nouns in the same noun phrase. On the other hand, in *baseline 3*, "*Condition 1*" is true only when the noun sequence do not have any modifier including clausal modifier and adjective modifier. Table 4 show the "*Condition 1*" for each baseline.

Table 4. *Condition 1* for each baseline.

	Condition 1
baseline 1	always true
baseline 2	true when the noun sequence is not modified by its preceding nouns in the same phrase
baseline 3	true when the noun sequence has no modifier

3.4 How to Use Synonym Knowledge

The basic strategy for determining a coreference relation is based on precise string matching between an anaphor and its antecedent candidate. We also make use of synonym knowledge to resolve a coreference relation that cannot be recognized by string matching.

In the synonym knowledge using methods, "*Condition 2*" is true not only when the anaphor exactly matches the antecedent candidate, but also when the anaphor is a synonym of the antecedent candidate.

3.5 How to Use Bridging Reference Resolution

We explain how to use the result of bridging reference resolution to coreference resolution. As mentioned, we do not consider a modified NP anaphoric in *baseline 2* and *baseline 3*. However, in some cases, an modified NP can be anaphoric. To deal with such cases, if two NPs share strings and have a bridging relation to the same entity, we consider the latter NP is anaphoric and has coreference relation to the former.

We use the method for bridging reference resolution proposed by Sasano et al.[9]. This method is based on automatically constructed nominal case frames. Nominal case frames are useful knowledge for resolving bridging reference and represents indispensable entities of the target noun.

(4) *Murayama shusho-wa nento-no kisha kaiken-de shokan-wo*
Murayama prime minister beginning of year press conference impressions
happyo-shita. Nento shokan-no yoshi-wa ika-no tori.
express beginning of year impressions point as follows

(Prime Minister Murayama expressed his impressions at the press conference of the beginning of the year. The point of the impressions is as follows.)

In example (4), the second "*shokan*"(impression) is modified by "*nento*"(beginning of year) and is not considered anaphoric in *baseline 2* or *baseline 3* method. However, "*shokan*"(impression) has a case frame named "AGENT" as shown in Table 5, and its bridging relation to "*shusho*"(prime minister) is recognized (i.e. the system recognize that the impression is the impression of the prime minister). Accordingly, the second "*shokan*" is considered anaphoric and the coreference relation between the first and the second "*shokan*" is recognized.

Table 5. Examples of nominal case frame.

Nominal case frame of "*shokan*"(impression)			
case frame	examples	:	frequency
AGENT	"*watashi*"(I)	:	24
	"*chiji*"(governor)	:	16
	"*sori*"(prime minister)	:	3
	"*hissha*"(writer)	:	2
	...	:	...

Nominal case frame of "*kekka*"(result)		
case frame	examples	: rate
"*koto*	"*chosa*"(investigation)	: 7648
(something)	"*senkyo*"(election)	: 1346
	"*enquête*"(questionnaire)	: 734
	"*jikken*"(experiment)	: 442
	...	: ...

"*koto*" = "*Aru koto-ga moto-ni natte okotta kotogara.*"
(a consequence, issue, or outcome of something)

In the methods using the result of bridging reference resolution, "*Condition 1*" is also true when the anaphor has a bridging relation, and then "*Condition 2*" is true only when the anaphor and it's antecedent candidate have the same referent of bridging.

As another example, although the second "*kekka*"(result) in example (5) is modified by "enquêtet" and is not considered anaphoric in *baseline 2* or *baseline 3* method, bridging reference resolver recognizes the two "*kekka*" refer to same entity "enquête" and the system recognizes the coreference relation between the first and the second "*result*".

(5) 2006 FIFA *world cup-no yushokoku yosou enquête-wo okonatta.*
2006 FIFA world cup winner expectation questionnaire conducted

Kekka-wa Brazil-ga top-datta. Kuwasii enquête kekka-wa HP-de.
result Brazil top detail questionnaire result web page

(The expectation questionnaire about 2006 FIFA world cup win ner was conducted. The top of the questionnaire result was Brazil. The detail of the result appeared in web page.)

4 Experiments

We conducted experiments on the Kyoto Corpus Version 4.0 [10]. In the corpus, coreference relations are manually annotated on the articles of *Mainichi* newspaper. We used 322 articles, which comprise 2098 sentences. These sentences have 2872 coreference tags that match our coreference criteria.

Table 6. Experimental results of coreference resolution.

method	precision	recall	F-score
baseline 1	57.0 (2246/3943)	**78.2** (2246/2872)	65.9
baseline 2	71.7 (2187/3052)	76.1 (2187/2872)	73.8
with bridging	71.5 (2200/3077)	76.6 (2200/2872)	74.0
with synonym	71.7 (2217/3092)	77.2 (2217/2872)	74.3
with syn. & brid.	71.5 (2231/3121)	77.7 (2231/2872)	**74.5**
baseline 3	77.4 (1966/2541)	68.5 (1966/2872)	72.6
with bridging	77.0 (1994/2590)	69.4 (1994/2872)	73.0
with synonym	**77.4** (1997/2581)	69.5 (1997/2872)	73.2
with syn. & brid.	77.0 (2025/2630)	70.5 (2025/2872)	73.6

We used 3 baseline methods, *baseline 1*, *baseline 2* and *baseline 3*. In addition, for *baseline 2* and *baseline 3*, we also conducted experiments with synonym knowledge and/or bridging reference resolution. Thus, all in all we conducted experiments in 9 different conditions.

$$F - score = \frac{2 \times Precision \times Recall}{Precision + Recall}. \qquad (a)$$

Table 6 shows the results of coreference resolution. F-score is calculated according to (a). *Baseline 1* achieve high recall but lowest precision and f-score. We can say that considering modified NPs as non-anaphoric improves F-score. We can also say that the condition used in *baseline 2*, "*Condition 1*" is true when the noun sequence is not modified by its preceding nouns in the same phrase, achieve best performance.

Furthermore, using knowledge of synonyms and the result of bridging reference resolution improves F-score and the usefulness of them is confirmed, but the effect is limited.

To investigate recall for several coreference types, we randomly selected 200 coreference tags from the Kyoto Corpus and evaluated the result of coreference resolution using *baseline 2* method with synonym knowledge and bridging reference resolution. Table 7 shows the recall for each coreference type.

The coreference relations that can be recognized by string matching are well recognized. On the other hand, the relations that need synonym knowledge to recognize are not (the recall is 50.0% (4/8)). However, 7 synonym relations out of 8 are included in automatically acquired knowledge of synonyms, and 3 coref-

Table 7. Recall for each coreference type.

relations between anaphor & antecedent	recall
1. anaphor's string is contained in antecedent's string	83.5 (142/170)
2. anaphor and its antecedent have a synonymous relation	50.0 (4/8)
3. other coreference types	0.0 (0/22)
sum	73.0 (146/200)

Table 8. Error analysis of erroneous system outputs.

error type	num
The anaphor and antecedent candidate refer to another entities	52
The possible anaphor is a general noun and not anaphoric	32
The antecedent candidate is a general noun and not anaphoric	7
others	9
sum	100

erence relations can not recognized only because the anaphors are modified. Therefore we can say that the coverage of the automatically acquired synonyms is not too small for resolving coreference relations between synonymous expressions.

The other types of coreference relations, such as relations between hypernym and hyponym, can not recognize fundamentally by our proposed method. To resolve such relations is our future work.

In order to investigate the cause of erroneous system outputs, we classify erroneous system outputs into 4 categories. Table 8 shows the classified error types of randomly selected 100 erroneous system outputs of *baseline 2* method with synonym knowledge and bridging reference resolution. Major erroneous system outputs were caused by two reasons:

1. *Baseline 2* method does not consider clausal or adjective modifiers.
2. Our system does not consider the generic usage of nouns.

In example (6), though the second "*jishin*"(earthquake) does not have coreference relation to "*Sanriku Harukaoki Jishin*", our system judges the two "*jishin*" refer to same entity because our system does not consider the modifiers "*yoshin-to mirareru*"(thought to be an aftershock).

Table 9. Comparison with previous work.

	precision	recall	F-score
Murata and Nagao	78.7 (89/113)	77.3 (89/115)	78.1
Iida et al.	76.7 (582/759)	65.9 (582/883)	70.9
Proposed	**71.5** (2231/3121)	**77.7** (2231/2872)	**74.5**

(6) *Sanriku Harukaoki Jishin-no yoshin-to mirareru jishin-ga hassei-shita.*
Far-off Sanriku Earthquake aftershock thought earthquake occurred

(An earthquake thought to be an aftershock of Far-off Sanriku Earthquake occurred.)

In example (7), although the second "wine" is used in generic usage, our system considers the second "wine" have coreference relation to "French wine" because our system does not consider generic usage of nouns.

(7) *Kare-wa* France-*no* wine-*ga suki-de kare-no ie-niwa* wine cellar-*ga aru.*
he French wine like his house wine cellar have

(He likes French wine and has wine cellar in his house.)

5 Related Work

Murata and Nagao proposed a rule-based coreference resolution method for determining the referents of noun phrases in Japanese sentences by using referential properties, modifiers and possessors [11]. As a result of experiments, they obtained a precision rate of 78.7% and a recall rate of 77.3%.

Their method performed relatively well. This may be because their experiments is constructed on small and supposedly easy corpus. Half of their corpus is occupied by fairy tale that is supposed to be easy to analyze.

Iida et al. proposed a machine learning approach for coreference resolution for Japanese [12]. Their process is similar to the model proposed by Ng and Cardie [13]. As a result of experiments on Japanese newspaper articles, they obtained a precision rate of 76.7% and a recall rate of 65.9%.

Table 9 shows the comparison with previous work and our proposed method. Since they used different data set and coreference criteria for experiments, these scores are not comparable as-is. However, taking into consideration Murata and Nagao uses small and supposedly easy corpus, we can say that our proposed method achieved enough performance.

Though these scores are not comparable as-is, rule-based methods outperformed learning-based methods in Japanese. This may be because recognizing most of coreference relations does not need complicated rules.

Bean and Riloff proposed a noun phrase coreference resolution system that uses information extraction patterns to identify contextual roles and creates four contextual role knowledge sources using unsupervised learning [14]. Experiments showed that the contextual role knowledge improved coreference performance for pronouns but not for noun phrases.

6 Conclusion

We have described a knowledge-rich approach to Japanese coreference resolution. We first proposed a method for acquiring knowledge of synonyms from large raw corpus and definition sentences of dictionaries for humans. Second, we proposed a method for improving coreference resolution by using the automatically acquired synonyms and the result of bridging reference resolution. Using the acquired synonyms and the result of bridging reference resolution boosted the performance of coreference resolution and the effectiveness of our integrated method is confirmed.

References

1. Yang, X., Zhou, G., Su, J., Tan, C.L.: Improving noun phrase coreference resolution by matching strings. In: Proceedings of 1st International Joint Conference of Natural Language Processing (IJCNLP04). (2004) 326–333
2. Soon, W.M., Ng, H.T., Lim, D.C.Y.: A machine learning approach to coreference resolution of noun phrases. Computational Linguistics **27**(4) (2001) 521–544
3. Ng, V.: Machine learning for coreference resolution: From local classification to global ranking. In: Proceedings of the 43rd Annual Meeting of the Asssociation for Computational Linguistics. (2005) 157–164
4. Luo, X., Ittycheriah, A., Jing, H., Kambhatla, N., Roukos, S.: A mention-synchronous coreference resolution algorithm based on the bell tree. In: Proceedings of the 42nd Meeting of the Association for Computational Linguistics. (2004) 135–142
5. Tajika, J., ed.: Reikai Syogaku Kokugojiten. Sanseido (1997)
6. Nishio, M., Iwabuti, E., Mizutani, S., eds.: Iwanami Kokugo Jiten. Iwanami Shoten (2000)
7. Kurohashi, S., Nagao, M.: A syntactic analysis method of long Japanese sentences based on the detection of conjunctive structures. Computational Linguistics **20**(4) (1994) 507–534
8. Isozaki, H., Kazawa, H.: Efficient support vector classifiers for named entity recognition. In: Proceedings of the 19th International Conference on Computational Linguistics. (2002) 390–396
9. Sasano, R., Kawahara, D., Kurohashi, S.: Automatic construction of nominal case frames and its applicatoin to indirect anaphora resolution. In: Proceedings of the 20th International Conference on Computational Linguistics. (2004) 1201–1207
10. Kawahara, D., Kurohashi, S., Hasida, K.: Construction of a Japanese Relevance-tagged Corpus. In: Proceedings of the 3rd International Conference on Language Resources and Evaluation. (2002) 2008–2013

11. Murata, M., Nagao, M.: An estimate of referent of noun phrases in Japanese sentences. In: Proceedings of 36th Annual Meeting of the Association for Computational Linguistics and the 17th International Conference on Computational Linguistics. (1998) 912–916
12. Iida, R., Inui, K., Matsumoto, Y., Sekine, S.: Noun phrase coreference resolution in japanese most likely candidate antecedents (in japanese). Journal of Information Processing Society of Japan **46**(3) (2005) 831–844
13. Ng, V., Cardie, C.: Improving machine learning approaches to coreference resolution. In: Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics. (2002) 104–111
14. Bean, D., Riloff, E.: Unsupervised learning of contextual role knowledge for coreference resolution. In: Proceedings of the Human Language Technology Conference / North American Chapter of the Assosication for Computational Linguistics Annual Meeting. (2004) 297–304