Japanese Predicate Argument Structure Analysis Exploiting Argument Position and Type

Yuta Hayashibe **Mamoru Komachi** **Yuji Matsumoto**
Graduate School of Information Science
Nara Institute of Science and Technology
8916-5, Takayama, Ikoma Nara 630-0192, Japan
{ yuta-h, komachi, matsu }@is.naist.jp

Abstract

We propose an approach to Japanese predicate argument structure analysis exploiting *argument position* and *type*. In particular, we propose the following two methods. First, in order to use information in the sentences in preceding context of the predicate more effectively, we propose an improved similarity measure between argument positions which is more robust than a previous co-reference-based measure. Second, we propose a flexible selection-and-classification approach which accounts for the minor types of arguments. Experimental results show that our proposed method achieves state-of-the-art accuracy for Japanese predicate argument structure analysis.

1 Introduction

The goal of predicate-argument structure analysis is to extract semantic relations such as "who did what to whom" that hold between a predicate and its arguments constituting a semantic unit of a sentence. It is an important step in many Natural Language Processing applications such as machine translation, summarization and information extraction.

Arguments are classified into three categories according to their positions relative to the predicates: intra-sentential arguments (those that have direct syntactic dependency with the predicates), zero intra-sentential arguments (those appearing as zero-pronouns but have their antecedents in the same sentence), and inter-sentential arguments (those appearing as zero-pronouns and their antecedents are not in the same sentence). We call them *INTRA_D, INTRA_Z*, and *INTER* respectively. Furthermore, we call these categories the *argument types*. While the analysis of INTRA_D is comparatively easy, INTRA_Z and INTER are more difficult. We consider that there are two reasons for this.

The first reason is the poverty of features for argument identification compared to INTRA_D. While for INTRA_D we have important clues such as the function word or directly dependency relation, we don't for INTRA_Z and INTER.

The second reason is the limited amount of training examples. For example, in a Japanese newswire corpus, INTRA_Z and INTER account for 30.5% and 12.4% of all the nominative (ga) cases, and 13.1% and 0.2% of all of the accusative (wo) cases (Iida et al., 2007).

In this paper, in order to solve these problems we propose the following two methods exploiting *argument position* and type.

First, we propose an improved similarity measure between argument positions of two predicates that take semantically similar arguments. For example, someone possibly arrested can also surrender him/herself, that is, objects of "arrest" and subjects of "surrender (oneself)" are occupied by semantically similar nouns. Gerber and Chai (2010) proposed analysis of English nominal predicates with this similarity to take discourse context into account. However, the similarity measure they used has drawbacks: it requires a co-reference resolver and a large number of documents. We improve their similarity measure alleviating these drawbacks by using argument position. We detail previous work on capturing discourse context in Section 2, and our proposal in Section 3.1.

Second, we propose a selection-and-classification approach. In this approach, in order to compensate for the relative infrequency of examples of INTRA_Z and INTER, we select a candidate argument for each argument type independently. After selecting candidates, we use classifiers to choose the correct argument type. This allows us to flexibly design features for each step and we can use pairwise features between the candidate arguments. We detail this in Section 3.2.

The experimental results demonstrated that our proposed method achieved the state-of-the-art of Japanese predicate argument structure analysis.

2 Related Work Capturing Discourse Context

2.1 Salient Reference List

Iida et al. (2003) used Salient Reference List (Nariyama, 2002) based on Centering Theory (Grosz et al., 1995), which explains the structure of discourse and the transition of topics in order to capture discourse context. The list has the following four ordered slots.

TOPIC (marked by *wa*-particle)
$>$ SUBJECT (*ga*)
$>$ INDIRECT OBJECT (*ni*)
$>$ DIRECT OBJECT (*wo*),

We check whether each candidate corresponds to any slots from the beginning of a document. If the candidate corresponds to a slot, we (over)write the slot with the candidate. We repeat this until we reach the predicate to analyze. We use the ranks of candidates in the list as a feature.

2.2 Argument Frequency

Iida et al. (2003) used a feature (*CHAIN LENGTH*) that stands for how often each candidate is used as an argument of predicates in preceding context. Imamura et al. (2009) used a similar binary feature (*USED*) that shows if each candidate is ever used as an argument of predicates or not. However, they did not investigate the effect of these features explicitly in their systems. Therefore we also investigate these in this paper.

2.3 Similarity between an Argument Position and a co-Reference Chain

In the study of implicit arguments[1] for English nominal predicates, Gerber and Chai (2010) used similarity features between an argument position and a co-reference chain, inspired by Chambers and Jurafsky (2008), who proposed unsupervised learning of narrative event chains using pointwise mutual information (PMI) between syntactic positions. This method stands on the assumption that similar argument positions tend to have the arguments which belong to a common co-reference chain. For instance, co-referring arguments at such argument positions like $\langle plead, ARG_0\rangle$, $\langle admit, ARG_0\rangle$, $\langle convict, ARG_1\rangle$, tend to take semantically similar nouns as the argument positions like $\langle sentence, ARG_1\rangle$, $\langle parole, ARG_1\rangle$.

They first automatically label a subset of the Gigaword corpus (Graff, 2003) with verbal and nominal semantic role labeling. They then identify co-references between arguments using a co-reference resolver. They compute PMI as follows.

Suppose the resulting data has N co-referential pairs of argument positions and M of these pairs comprising $E_a = \langle P_a, A_a\rangle$, $E_b = \langle P_b, A_b\rangle$, and $E_c = \langle P_c, A_c\rangle$. P_a, P_b, and P_c are predicates, and A_a, A_b, and A_c are labels such as ARG_0 or ARG_1.

$$pmi(E_a, E_b) = \log \frac{G(E_a, E_b)}{G(E_a, *)G(E_b, *)}$$

$$G(E_a, E_b) = \frac{M}{N}$$

With this similarity between argument positions, they defined scores between an argument position and a co-reference chain.

3 Predicate Argument Structure Analysis Exploiting Argument Position and Type

3.1 Similarity between Argument Positions using Distribution Similarity

Suppose we want to identify the argument of **自首した** (surrendered) in Example (1). The argument is an antecedent of zero-pronoun ϕ of the predicate.

	police	*wa*-particle	hanako	*wo*-particle	arrested
(1)	警察	は	花子	を	逮捕した.

Police arrested Hanako.

I		had surrendered	that	heard
私は	(ϕ が)	**自首した**	と	聞いた.

I heard that ϕ *had surrendered*.

With Salient Reference List for “自首する (surrendered)”, the rank of “警察 (police)” is higher than that of “花子 (Hanako)” and it is noisy information for analysis. We also cannot distinguish them with argument frequency information, because frequencies of both “花子 (Hanako)” and “警察 (police)” are 1.

Though it is reasonable to use the similarity between an argument position and a co-reference

[1] In short, this is equivalent to INTER.

1599	が[ga] 自首する (surrender)	5651	が[ga] 逮捕する (arrest)	82112	を[wo] 逮捕する (arrest)
136	-者 person	702	署員 PS staff	15285	人 person
117	犯人 criminal	698	警察 police	8484	-者 person
96	彼 he	376	警察官 police officer	5563	男 man
68	人 person	368	-署 police station	2804	-名 name
63	男 man	230	県警 prefectural police	1188	犯人 criminal
36	-犯 crime	177	-員 -staff	1185	男性 male
30	少年 boy	153	府警 prefectural police	763	-ら -s
26	高校生 high-scool student	137	当局 authorities	671	おまえ you

Table 1: Argument distributions of each argument position (Sorted by frequency)

chain, the similarity measure described in Section 2.3 has two problems.

One is the strong dependency on the accuracy of co-reference resolver system. In fact, the accuracy of Japanese co-reference resolvers is not accurate enough to create co-reference chains in good quality.[2] The other problem is the problem that it needs a lot of documents, because the method does not use any non co-referring nouns.

To avoid using an unreliable co-reference resolver, we can suppose the same noun lemmas without pronouns in the same document are co-references. Pekar (2006) called the noun lemmas *anchors* and they supposed the similarity measure between syntactic positions. For example, there are two anchors: "Mary" and "house" in the sentences "Mary bought a house. The house belongs to Mary." They extract two groups: { buy(obj:X), belong(subj:X) } and {buy(subj:X), belong(to:X). } Nevertheless, this method also requires many documents because noun lemmas without anchors are not used for the calculation.

In this paper, we propose a more robust similarity measure between argument positions which does not depend on unreliable co-reference annotations by the resolver.

Table 1 shows the list of nouns that have direct dependency arcs in syntactic dependency structures along with case markers が[ga] (nominative case), を[wo] (accusative case) and に[ni] (dative case) extracted from the *WEB* corpus described in Section 4. According to Table 1, the distributions of nouns of "自首する (surrender)" following が[ga] and "逮捕する (arrest)" following を[wo] look alike. We can expect that an arrested person is more likely to be a person who has surrendered than an arrestee. We define a novel similarity of two argument positions

[2] We implemented the method proposed by Iida et al. (2005a), and the F-measure was 66%.

		E_1	E_2	E_3
E_1	が[ga] 自首する (nominative case of *surrender*)	0	0.5409	0.2431
E_2	が[ga] 逮捕する (nominative case of *arrest*)	0.5409	0	0.5139
E_3	を[wo] 逮捕する (accusative case of *arrest*)	0.2431	0.5139	0

Table 2: An example of similarities between argument positions calculated with *WEB* corpus.

encoding such information as Jensen-Shannon divergence between argument distributions of argument positions. A sample of the calculated similarities is shown in Table 2. This table illustrates the most similar argument positions are the nominative (が[ga]) case of "自首する (surrender)" (E_1) and the accusative (を[wo]) case of "逮捕する (arrest)" (E_3). We will use these values as features of predicate-argument analysis in the experiments.

3.2 Selection-and-Classification Approach Considering Argument Type

In previous work, argument analysis was performed with common features regardless its argument type. However, these methods have difficulty in distinguishing the marginal cases where two candidates have different argument types because of the difference of quantity by argument types. Thus we propose the *selection-and-classification approach* for Japanese predicate argument structure analysis. This approach consists of two steps: the *selection* step and the *classification* step.

This approach is inspired by two models. The first is the selection-and-classification model (Iida et al., 2005b) for noun phrase anaphora resolution. The model first selects a likely antecedent of the target (possibly) anaphoric expression. Second, the model classifies the target anaphoric ex-

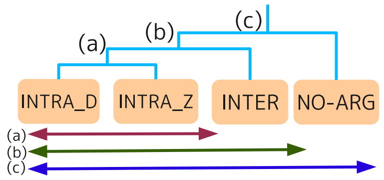

Figure 1: Argument selection in the ‘Classification’ step from three most likely argument candidates

	Label	Arguments to make features
a	INTRA_D	⟨*INTRA_D*, INTRA_Z⟩
	INTRA_Z	⟨INTRA_D, *INTRA_Z*⟩
b	INTRA	⟨*INTRA_D*, INTER⟩, ⟨*INTRA_Z*, INTER⟩
	INTER	⟨{INTRA_D or INTRA_Z }, *INTER*⟩
c	HAVE-ARG	⟨{INTRA_D or INTRA_Z or INTER}⟩
	NO-ARG	⟨*INTRA_D*⟩, ⟨*INTRA_Z*⟩, ⟨*INTER*⟩

Table 3: Examples made for training. *Italic texts* in (b) and (c) refer most likely argument which (a) and (b) selected respectively. Non-italic texts refer to the correct argument.

pression either true anaphoric or not with the most likely antecedent. They took this approach since there are almost no clues for a Japanese noun phrase to determine anaphoric or not by looking only at the noun phrase.

Similarly, in our approach, after selecting the most likely candidate of the argument, we classify it into either of INTRA_D, INTRA_Z, INTER or no argument.

The second is the *tournament model* (Iida et al., 2003) for zero-anaphora resolution. For all the candidate antecedents (virtually all noun phrases appearing in preceding context), the model repeats two-class classification: which candidate in the pair of candidates is likely to be the antecedent for the zero-anaphora. The advantage of the tournament model is that the model can use pairwise features of candidates. Similarly, in the classification step of our approach we select an argument comparing most likely candidates of arguments of each argument type.

A Method of Argument Analysis

Selection: At the first step, we select three most likely arguments of INTRA_D, INTRA_Z, and INTER for each predicate using any argument identification model. We may use different features for models of different argument types.

Classification: At the second step, we determine which INTRA_D, INTRA_Z, and INTER is the correct argument or if there is no explicit argument appearing in the context. This step is composed

Gold Scope	Example made for training
NO-ARG	(c)NO-ARG
INTER	(b)INTER, (c)HAVE-ARG
INTRA_Z	(a)INTRA_Z, (b)INTRA, (c)HAVE-ARG
INTRA_D	(a)INTRA_D, (b)INTRA, (c)HAVE-ARG

Table 4: Examples made for training with human-annotated data. Generated examples depend on the argument type.

of three binary classification models illustrated in Figure 1.

(a) Judge which of INTRA_D or INTRA_Z is more likely to be an argument of the predicate.
(b) Judge which of INTER or the candidate selected at (a) is more likely to be an argument of the predicate.
(c) Judge whether the candidate selected at (b) qualifies as an argument of the predicate or not.

We show the example of analysis of ϕ in Example (1). We first select argument candidates of INTRA_D, INTRA_Z, and INTER. Suppose the most likely argument of INTRA_D is not selected and “私 (I)” and “花子 (Hanako)” are selected as ones of INTRA_Z and INTER respectively in the ‘selection’ step. Because INTRA_D is not selected, the classifier selects INTRA_Z at (a). Suppose INTER is selected at (b) comparing “私” selected at (a) and “花子”. Finally, “花子” is selected as the argument by the classifier of (c).

Furthermore, though we tried different orders for ‘Classification’ step in the preliminary experiment, this order was the best.

Training Method of Classifiers for the ‘Classification’ Step

We train each binary classifier in the order of (a), (b), and (c). We create training examples of classifiers with two argument candidates and a predicate as shown in Tables 3 and 4. The following arguments are used for training:

(a) the correct argument and the most likely argument selected at the ‘Selection’ step
(b) the correct argument and the most likely argument selected by (a) at the ‘Classification’ step
(c) the correct argument and the most likely argument selected by (b) in the ‘Classification’ step

For instance, ϕ in Example (1) is “花子 (Hanako)” whose argument type is INTER. Hence

CHAIN_LENGTH	A frequency of being arguments in previous sentences (Iida et al., 2003)
USED	Whether being arguments in former sentences or not (Imamura et al., 2009)
SIM_COREF	Scores between an argument position and co-reference chain calculated with the similarity which Gerber and Chai (2010) used (described in Section 2.3)
SIM_CS	Scores between an argument position and co-reference chain calculated with our proposed similarity

Table 5: Discourse context features used in the experiment

we generate two training examples: One is an example of (b) with the label INTER, “花子”, and the most likely argument selected by (a) at ‘Classification’ step. The other one is an example of (c) with the label HAVE-ARG and “花子”.

4 Evaluation Setting of Predicate Argument Structure Analysis Exploiting Argument Position and Type

We evaluate our proposed selection-and-classification approach by comparing it with other models and the discourse context features shown in Table 5 by adding them to the baseline features at Japanese predicate argument structure analysis of nominative case.

In the experiment, systems refer only nouns in co-reference chains which are intra-sentential arguments. In addition, we used human annotated data of co-reference and predicate-argument structure to make discourse context features. For SIM_COREF and SIM_CS, we used maximum, minimum and average scores of similarities.

4.1 Dataset for Similarity Calculation

We used two datasets for the calculation of similarities: the Newspapers (NEWS) and the Web texts (WEB).

NEWS: We used about 21,000,000 sentences in Mainichi newspapers published from 1991 to 2003 (excluded 1995). We part-of-speech tagged the data with MeCab 0.98[3] and dependency structure parsed with CaboCha 0.60pre4[4]. Both taggers used the NAIST Japanese Dictionary 0.6.3[5]. We extracted 27,282,277 pairs of a predicate and an argument.[6]

We also extracted 111,173,873,092 co-reference chains to calculate SIM_COREF with the anaphora resolver which is our re-implementation of (Iida et al., 2005a). These chains include 2,280,417,516,455 nouns. We used 173,778,624 pairs of a predicate and an argument with the case maker が(ga), を(wo) and に(ni).

WEB: We used about 500,000,000 sentences which Kawahara and Kurohashi (2006) collected from the web. They are part-of-speech tagged with JUMAN[7] and dependency structure parsed with KNP[8]. We extracted 1,101,472,855 pairs of a predicate and an argument.[9]

4.2 Training and Evaluation Dataset

We used NAIST Text Corpus 1.4β (Iida et al., 2007) for training and evaluation. It is based on Kyoto Text Corpus 3.0[10] and annotated with predicate-argument structure, event noun structure, and co-reference of nouns about 40,000 sentences of Japanese newspaper text. We excluded 11 articles due to annotation error.

We conducted five-fold cross-validation. In the experiments, base phrases and dependency relations are acquired from the Kyoto Text Corpus 3.0 in the same way of related work.

4.3 A Model in the ‘Selection’ Step

In order to identify the most likely argument candidate of each INTRA_D, INTRA_Z, and INTER, we used the tournament model. We emphasize that our proposed approach can use any argument identification model to identify the most likely candidate of an argument.

4.4 Baseline Features and Classifier

As baseline features, we employed features proposed by Iida et al. (2005a, 2007a) and Imamura et al. (2009) in addition to a novel one ‘PRED DEP POS’ shown in Table 6.

We used Support Vector Machine (Cortes and Vapnik, 1995) for each classification model with

[3] http://mecab.sourceforge.net/

[4] http://chasen.org/~taku/software/cabocha/

[5] http://sourceforge.jp/projects/naist-jdic/

[6] Unique total are; Verb: 31,012, Noun: 327,174, Pair: 7,071,627

[7] http://nlp.kuee.kyoto-u.ac.jp/nl-resource/juman.html

[8] http://nlp.kuee.kyoto-u.ac.jp/nl-resource/knp.html

[9] Unique total are; Verb:about 801 million, Noun: about 288 million, Pair: 15,994 million

[10] http://nlp.kuee.kyoto-u.ac.jp/nl-resource/corpus.html

Type	Name	Description (NP and PRED refer a noun phrase and a predicate.)
Lexical	HEAD BF	Head word of NP and PRED.
Grammatical	PRED IN MATRIX	1 if PRED exists in the matrix clause; otherwise 0.
	PRED IN EMBEDDED	1 if PRED exists in the relative clause; otherwise 0.
	PRED VOICE	1 if PRED contains auxiliaries such as '(ra)reru'; otherwise 0.
	PRED AUX	1 if PRED contains auxiliaries such as '(sa)seru', 'hosii', 'morau'
	PRED ALT	0 if PRED VOICE and PRED AUX are 1; otherwise 1.
	POS	Part-of-speech of NP.
	DEFINITE	1 if NP contains the article corresponding to DEFINITE 'the', such as 'sore' or 'sono'; otherwise 0.
	DEMONSTRATIVE	1 if NP contains the article corresponding to DEMONSTRATIVE 'that' or 'this', such as 'kono', 'ano'; otherwise 0.
	PARTICLE	Particle followed by NP.
	PARTICLE IF PRED VOICE	PARTICLE followed by NP if PRED VOICE is 1.
Semantic	NE	Named entity of NP: PERSON, ORGANIZATION, LOCATION, ARTIFACT, DATE, TIME, MONEY, PERCENT or N/A.
	PRONOUN TYPE	Pronoun type of NP. (e.g. 'kare (he)': PERSON, 'koko (here)': LOCATION, 'sore (this)': OTHERS)
	SELECT REST	1 if NP satisfies selectional restrictions in Nihongo Goi Taikei (Japanese Lexicon) (Ikehara et al., 1997); otherwise 0.
	NCV NPS PMI	PMI score of [⟨Noun, Case⟩,⟨Predicate⟩] calculated from the WEB corpus.
Positional	DIST NP PRED	Distance between NP and PRED.
	DIST NPS	Distance between NP and NP.
	BEGINNING	1 if NP is located in the beginning of sentence; otherwise 0.
	END	1 if NP is located in the end of sentence; otherwise 0.
	PRED NP	1 if PRED proceeds NP; otherwise 0.
	NP PRED	1 if NP precedes PRED; otherwise 0.
	DEP PRED	1 if NP depends on PRED; otherwise 0.
	DEP NP	1 if PRED depends on NP; otherwise 0.
	PRED DEP POS	Part-of-speech of head word depended by PRED.
	IN QUOTE	1 if NP exists in the quoted text; otherwise 0.
	PATH	Dependency relation between PRED and NP.
Context	SRL RANK	A rank of NP in Salient Reference List.
	GA REF	1 if the phrase which contains noun and 'ga' depend on NP with 'nagara', 'te', 'shi', 'tsutsu', 'tameni', 'tari', ϕ; otherwise 0.

Table 6: Baseline features of classifiers in the 'Selection' step and the 'Classification' step.

a linear kernel. We used the implementation of LIBLINEAR 1.7[11] with its default parameters.

4.5 Targets for Comparison of Predicate Argument Analysis Model

We evaluate our selection-and-classification approach by comparing our baseline model with two previous approaches TA and IM.

TA: Taira et al. (2008) used decision lists where features were sorted by their weights learned from Support Vector Machine. They simultaneously solved the argument of event nouns in the same lists.

IM: Imamura et al. (2009) used discriminative models based on maximum entropy. They added the special noun phrase NULL, which expresses that the predicate does not have any argument.

Because previous work use different features and machine learning methods and experiment on different setting from ours, we also compare with a baseline model BL in order to analyze the effect of dividing a model considering argument type.

BL: This model has a single step in 'classification' step. In other words, 'selection' step in this model selects the most likely argument from all noun phrases preceding the predicate.

5 Discussion

Table 7 presents the result of the experiments. According to the bottom row in Table 7, we achieved the state-of-the-art of Japanese predicate argument structure analysis by combining all discourse context features (+A+B+C+D+E).

We investigate our result from five different standpoints.

5.1 Effect of the Selection-and-Classification Approach

We analyze the effect of our proposed selection-and-classification approach by comparing the first row of Table 7. SC is superior to BL in all types. This shows that dividing a model considering argument type improves the performance.

[11] http://www.csie.ntu.edu.tw/~cjlin/liblinear/

Section		INTRA_D			INTRA_Z			INTER		
		P	R	F_1	P	R	F_1	P	R	F_1
5.1	BL : Our baseline	84.06	50.74	63.24	27.02	56.13	36.46	16.44	13.70	14.89
	SC : Our proposed method	80.71	85.35	82.96	47.57	45.74	46.64	23.79	15.93	19.07
5.2	TA : Taira et al. 2008	-	-	75.53	-	-	30.15	-	-	23.45
	IM : Imamura et al. 2009	85.2	88.8	87.0	58.8	43.4	50.0	47.5	7.6	13.1
5.3, 5.4	SC+A (CHAIN_LENGTH)	85.39	88.79	87.05	51.64	52.22	51.93	25.31	18.63	21.44
	SC+B (USED)	85.59	88.44	86.99	54.40	53.32	53.86	**26.09**	21.27	**23.43**
	SC+C (SIM_COREF_NEWS)	86.82	88.90	87.85	54.07	52.89	53.47	25.83	20.08	22.58
	SC+D (SIM_CS_NEWS)	**88.42**	**91.10**	**89.74**	59.05	58.12	58.58	24.81	19.91	22.08
	SC+E (SIM_CS_WEB)	87.00	90.44	88.69	**64.76**	**60.27**	**62.43**	25.63	**21.32**	23.27
5.4	SC+D+E	89.70	91.55	90.62	65.08	61.37	63.17	24.86	21.57	23.08
5.5	ALL (SC+A+B+C+E+D)	89.93	91.70	90.81	67.39	62.18	64.68	25.86	22.93	24.30
	ALL-A	90.44	91.34	90.89	68.12	61.95	64.89	25.72	23.77	24.69
	ALL-B	90.48	91.48	90.98	66.83	62.06	64.35	**25.48**	22.71	24.01
	ALL-C	89.30	91.65	90.46	65.19	61.92	63.50	25.71	**22.22**	**23.83**
	ALL-D	**87.65**	**90.48**	**89.04**	66.14	61.21	63.57	26.03	22.85	24.32
	ALL-E	89.66	90.73	90.19	**62.47**	**59.04**	**60.71**	25.56	22.49	23.92

Table 7: Comparison of predicate argument structure analysis of nominative case: P, R, and F_1 indicate Precision, Recall, and F-measure($\beta = 1$), respectively.

5.2 Comparison between previous work

By comparing SC and TA, and SC+USED and IM[12], the result of our proposed method is competitive or superior to others. Additionally, recall is higher in any type; therefore we consider there is still much room for improvement by replacing the argument identification model in the selectional step with other models.

5.3 Effect of Similarity Metrics

On comparing +A (CHAIN_LENGTH), +B (USED), and +C (SIM_COREF_NEWS) or +D (SIM_CS_NEWS) in Table 7, similarity-based features are superior or competitive to frequency-based feature.

(2) 婚姻は毎年一万—四万組も増え、 …
The number of marriages increases 10,000 to 40,000 couples annually …

… **流行して** … の引き金となったインフルエンザが、
The flu that *has been going around* and triggered …,

For instance, the argument of “流行し (be going around)” in Example (2) is “インフルエンザ (flu)” of INTER and is not an argument of previous arguments. Though the topic changes between two sentences, A and B cannot take it into account this and output “婚姻 (Marriages)” which is an argument of “増え (increase)” because the frequency-based feature is active. In contrast, C and D handle this because the similarity between the nominative case of “増え” and “流行し” is low.

On comparing +C (SIM_COREF_NEWS) and +D (SIM_CS_NEWS) in Table 7, our proposed similarity metrics work better than the co-reference-based metrics in INTRA_D or INTRA_Z by a large margin. This result shows the robustness of our metrics compared to the co-reference based similarity between argument positions.

5.4 Effect of In and Out-of-domain Data

On comparing +D (SIM_CS_NEWS) and +E (SIM_CS_WEB) in Table 7 respectively, the similarity measure using the newswire texts works better for INTRA_D and one using the web texts works better for INTRA_Z and INTER.

Additionally, the result of +D+E shows that combining proposed similarities calculated from different sources work complementary.

5.5 Ablation Features

Removing features one by one from ALL (Adding all of A to E), we inquire about features which have strong effect on ALL. Table 7 shows that the F-measures of INTRA_D and INTRA_Z fall by a large margin, by removing D and E respectively. Though the F-measure of INTER degrades by removing C, it makes little difference to other argument types. This shows it is our proposed similarity that mainly contributes to the improvement of the F-measure of the overall system.

6 Error Analysis

We analyze errors where our proposed similarity does not work well.

[12] We compare SC+USED and IM, because IM used the USED feature.

6.1 Copula

In NAIST Text Corpus, the copula "だ"(In English, "be") is annotated as a predicate.

(3) マンション価格が ... 下がってきた ...。
The price of apartments is going down.

...（ϕ が） 前年は五・八倍**で**あった
ϕ of last year *was* 5.8 times higher than that of this year.

However, the behavior of copula is different from other predicates, thus it is difficult to resolve them with the same model. To solve this problem, the model of copula should be separated.

6.2 Light Verb Construction

In this experiment, we regarded the predicate of "する (do) + noun" such as "逮捕する (do arrest)" as a single predicate. On the other hand, we regarded the predicate of "する (do) + particle + noun" such as "まかせにする" in Example (4) as "する (do)."

(4) （ϕ が） 分権を国 まかせ (relegation) に (ni) **する** *(do)* のではなく、...
ϕ should not relegate promotion of decentralization ...

However, verbs like "do" in such examples do not play central roles, whereas the noun such as "まかせ (relegation)" carries the main meaning of the event. This phenomenon is called "light verb construction" (Miyamoto, 1999).

"まかせ" is the nominalized form of the verb "まかせる (relegate)." Thus we need to calculate similarity with "まかせる" instead of "する". When the predicate is a light verb, we have to use the original verb to calculate the similarity.

6.3 Predicate Sense Ambiguity

A predicate may have several senses and hence have several argument distributions. For example, "詰める" has two senses at least: to pack and to bring to a conclusion. "詰める" in Example (5) means the latter.

(5) （ϕ が） 数字を早急に**詰める**必要性を強調した。
They emphasized that ϕ should be brought to an conclusion as soon as possible.

The distributions of arguments of such ambiguous verbs tend to have a mixture of several distributions of arguments. Therefore this makes it hard to calculate the similarity of an argument position and a co-reference chain. Additionally, it is even more difficult when the predicate is more essential verb such as "持つ (have)" and "取る (take)." This suggests the close relationship between the word sense disambiguation and the predicate argument structure analysis. In fact, Meza-Ruiz and Riedel (2009) showed that the joint model for semantic role labeling and word sense disambiguation performs better than a pipeline system.

Since NAIST Text Corpus is not annotated with verb senses, we are annotating the sense of verbs to allow similar analysis.

7 Conclusion

We improved Japanese predicate argument structure analysis exploiting *argument position* and *type*. In particular, we proposed two methods: the improved similarity measure between argument positions and the selection-and-classification approach considering argument type.

Experimental results show that our proposed method achieved state-of-the art accuracy for the Japanese predicate argument structure analysis. Proposed similarity between argument positions exploiting case maker is more robust than previous co-reference-based method that makes use of an unreliable automatic co-reference resolver. Furthermore, we proposed flexible approach which accounts for the minor types of arguments.

Future work includes four topics: (1) to distinguish copula from other predicates; (2) to combine internal argument to take semantic argument into consideration if the verb is in light verb construction; (3) to perform word sense disambiguation before calculating similarity; (4) to conduct experiments not only on nominative case but also on other cases .

Acknowledgments

We thank Daisuke Kawahara and Sadao Kurohashi for providing the web texts and Joseph Irwin for his comments on the earlier draft.

References

Nathanael Chambers and Dan Jurafsky. 2008. Unsupervised Learning of Narrative Event Chains. In

Proceedings of Annual Meeting of the Association for Computational Linguistics-08 with the Human Language Technology Conference, pages 789–797.

Corinna Cortes and Vladimir Vapnik. 1995. Support-Vector Networks. *Machine learning*.

Matthew Gerber and Joyce Y. Chai. 2010. Beyond NomBank: A Study of Implicit Arguments for Nominal Predicates. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 1583–1592. Association for Computational Linguistics.

David Graff. 2003. *English Gigaword*. Linguistic Data Consortium.

Barbara J. Grosz, Aravind K. Joshi, and Scott Weinstein. 1995. Centering: A Framework for Modeling the Local Coherence of Discourse. *Computational Linguistics*, 21(2):203–225.

Ryu Iida, Kentaro Inui, Hiroya Takamura, and Yuji Matsumoto. 2003. Incorporating Contextual Cues in Trainable Models for Coreference Resolution. In *Proceedings of the 10th EACL Workshop on the Computational Treatment of Anaphora*, pages 23–30.

Ryu Iida, Kentaro Inui, and Yuji Matsumoto. 2005a. Anaphora Resolution by Antecedent Identification Followed by Anaphoricity Determination. *ACM Transactions on Asian Language Information Processing*, 4(4):417–434.

Ryu Iida, Kentaro Inui, and Yuji Matsumoto. 2005b. On the Issue of Combining Anaphoricity Determination and Antecedent Identification in Anaphora Resolution. In *International Conference on Natural Language Processing and Knowledge Engineering*, pages 244–249.

Ryu Iida, Mamoru Komachi, Kentaro Inui, and Yuji Matsumoto. 2007. Annotating a Japanese Text Corpus with Predicate-Argument and Coreference Relations. In *Proceedings of the Linguistic Annotation Workshop*, pages 132–139. Association for Computational Linguistics.

Satoru Ikehara, Masahiro Miyazaki, Satoshi Shirai, Akio Yokoo, Hiromi Nakaiwa, and Kentaro Ogura. 1997. *Nihongo Goi Taikei :A Japanese Lexicon*. Iwanami Shoten.

Kenji Imamura, Kuniko Saito, and Tomoko Izumi. 2009. Discriminative Approach to Predicate-Argument Structure Analysis with Zero-Anaphora Resolution. In *Proceedings of the Joint conference of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing*, pages 85–88.

Daisuke Kawahara and Sadao Kurohashi. 2006. Case Frame Compilation from the Web using High-Performance Computing. In *Proceedings of the 5th International Conference on Language Resources and Evaluation*, pages 1344–1347.

Ivan Meza-Ruiz and Sebastian Riedel. 2009. Jointly identifying predicates, arguments and senses using markov logic. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 155–163.

Tadao Miyamoto. 1999. *The Light Verb Construction in Japanese: the role of the verbal noun*. John Benjamins Publishing Company.

Sigeko Nariyama. 2002. Grammar for Ellipsis Resolution in Japanese. In *Proceedings of the 9th International Conference on Theoretical and Methodological Issues in Machine Translation*, pages 135–145.

Viktor Pekar. 2006. Acquisition of Verb Entailment from Text. In *Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics*, pages 49–56. Association for Computational Linguistics.

Hirotoshi Taira, Sanae Fujita, and Masaaki Nagata. 2008. A Japanese Predicate Argument Structure Analysis Using Decision Lists. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 523–532.