

Predicate-Argument Structure Analysis with Zero-Anaphora Resolution for Dialogue Systems

Kenji Imamura, Ryuichiro Higashinaka, and Tomoko Izumi

NTT Media Intelligence Laboratories, NTT Corporation

1-1 Hikari-no-oka, Yokosuka, 239-0847, Japan

{imamura.kenji,higashinaka.ryuichiro,izumi.tomoko}@lab.ntt.co.jp

Abstract

This paper presents predicate-argument structure analysis (PASA) for dialogue systems in Japanese. Conventional PASA and semantic role labeling have been applied to newspaper articles. Because pronominalization and ellipses frequently appear in dialogues, we base our PASA on a strategy that simultaneously resolves zero-anaphora and adapt it to dialogues. By incorporating parameter adaptation and automatically acquiring knowledge from large text corpora, we achieve a PASA specialized to dialogues that has higher accuracy than that for newspaper articles.

1 Introduction

Semantic role labeling (SRL) and predicate-argument structure analysis (PASA) are important analysis techniques for acquiring “who did what to whom” from sentences¹. These analyses have been applied to written texts because most annotated corpora comprise newspaper articles (Carreras and Màrquez, 2004; Carreras and Màrquez, 2005; Matsubayashi et al., 2014).

Recently, systems for speech dialogue between humans and computers (e.g., Siri of Apple Inc. and Shabette Concier of NTT DoCoMo) have become familiar with the popularization of smart phones. A man-machine dialogue system has to interpret human utterances to associate them with system utterances. The predicate-argument structure could be an effective data structure for dialogue management. However, it is unclear whether we can apply the SRL/PASA for newspaper articles to dialogues because there are many differences between them, such as the number of speakers, written or spoken language, and context processing. For example, the following dialogue naturally includes pronouns, and thus anaphora resolution is necessary for semantic role labeling.

A:	[I] _{ARG0}	want	[an iPad Air] _{ARG1} .
B:	[When] _{ARGM}	will	[you] _{ARG0} buy [it(=an iPad Air)] _{ARG1} ?

Similar phenomena exist in Japanese dialogues. However, most pronouns are omitted (called zero-pronouns), and zero-anaphora resolution is necessary for Japanese PASA.

A:	[iPad Air] _{NOM}	-ga	<i>hoshii-na.</i>
	iPad Air	NOM.	want
	“ ϕ want an iPad Air.”		
B:	<i>itsu</i>	ϕ _{NOM} ϕ _{ACC}	<i>kau-no?</i>
	when		buy?
	“When will ϕ buy ϕ ?”		

This work is licensed under a Creative Commons Attribution 4.0 International License. Page numbers and proceedings footer are added by the organizers. License details: <http://creativecommons.org/licenses/by/4.0/>

¹Recent SRL systems assign labels of predicates and their arguments as semantic roles. Consequently, SRL and PASA are very similar tasks. We use the term predicate-argument structure analysis in this paper because most Japanese analyzers use this term.

This paper presents predicate-argument structure analysis with zero-anaphora resolution for Japanese chat dialogues. Here, we regard the task of constructing PASA for dialogues as a kind of domain adaptation from newspaper articles to dialogues. Màrquez et al. (2008) and Pradhan et al. (2008) indicated that the tuning of parameter distribution and reducing the out-of-vocabulary are important for the domain adaptation of SRL. We also focus on parameter distribution and out-of-vocabulary to construct a PASA adapted to dialogues. To the best of our knowledge, this is the first paper to describe a PASA for dialogues that include many zero-pronouns.

The paper is organized as follows. Section 2 briefly reviews SRL/PASA in English and Japanese. Section 3 discusses characteristics of chat dialogues by comparing two annotated corpora, newspaper articles and dialogues. Section 4 describes the basic strategy of our PASA, and Section 5 shows how it was adapted for dialogues. Experiments are presented in Section 6, and Section 7 concludes the paper.

2 Related Work

2.1 Semantic Role Labeling in English

The advent of the supervised method proposed by Gildea and Jurafsky (2002) has led to the creation of annotated corpora for semantic role labeling. In the CoNLL-2004 and 2005 shared task (Carreras and Màrquez, 2004; Carreras and Màrquez, 2005), evaluations were carried out using the Proposition Bank (Palmer et al., 2005). Because the Proposition Bank was annotated to the Penn Treebank (i.e., the source texts were from the Wall Street Journal), the shared tasks were evaluated on newspaper articles. Màrquez et al. (2008) provides a review of SRL.

OntoNotes Corpus (Hovy et al., 2006) contains multiple genres such as newswire, broadcast news, broadcast conversation. The annotation to OntoNotes includes semantic role labels compliant with the Proposition Bank. It is currently used for coreference resolution (Pradhan et al., 2012), and is expected to be applied to dialogue analysis.

A few SRL studies have focused on not only verbal predicates (e.g., ‘*decide*’) but also nominal predicates (e.g., ‘*decision*’) (Jiang and Ng, 2006; Gerber and Chai, 2012; Laparra and Rigau, 2013). Because the subject and object of nominal predicates are frequently omitted (e.g., the object in the phrase “*the decision*” is omitted), problems similar to the Japanese zero-pronouns have to be resolved in the SRL of nominal predicates.

2.2 Predicate-Argument Structure Analyses in Japanese

Japanese material includes the NAIST Text Corpus (Iida et al., 2007)², which is an annotated corpus of predicate-argument structures and coreference information for newspaper articles. Argument noun phrases of the nominative, accusative, and dative cases are assigned to each predicate. The predicate and the noun phrases are not limited to the same sentence. If arguments of the predicate are represented as zero-pronouns, the antecedent noun phrases in other sentences are assigned as the arguments.

Many PASA methods have been studied on the NAIST Text Corpus (Komachi et al., 2007; Taira et al., 2008; Imamura et al., 2009; Yoshikawa et al., 2011). In Japanese, some of them simultaneously resolve the zero-anaphora caused by zero-pronouns.

Most English SRL and Japanese PASA currently target newspaper articles, and it is unclear whether the methods for newspapers can be applied to dialogue conversations.

3 Characteristics of Chat Dialogues

We first collected chat dialogues of two speakers and annotated them with the predicate-argument structure. The participants chatted via keyboard input. Therefore, fillers and repetitions, which are frequent in speech dialogues, were rare. The theme was one of 20 topics, such as meals, travel, hobbies, and TV/radio programs. Annotation of the predicate-argument structure complied with the NAIST Text Corpus. Figure 1 shows a chat dialogue example and its predicate-argument structure annotation.

²<http://cl.naist.jp/nldata/corpus/>. We use version 1.5 with our own preprocessing in this paper. NAIST is an acronym of “Nara Institute of Science and Technology.”

A:	<i>natsu-wa</i> (exo2) _{NOM} (exog) _{DAT} dekake-tari-shimashi-ta-ka? “Did (you) _{NOM} go (anywhere) _{DAT} in this summer?”
B:	<i>8-gatsu-wa Ito-no</i> [<i>hanabi-taikai</i>] _{DAT-ni} (exo1) _{NOM} yuki-mashi-ta. “(I) _{NOM} went to [the fireworks*1] _{DAT} at Ito in August.”
A:	[<i>hanabi</i> *2] _{ACC} , [<i>watashi</i> *3] _{NOM} - <i>mo mi-takatta-desu.</i> “[Fireworks*2] _{ACC} , [I*3] _{NOM} also wanted to see (it).”
A:	<i>demo, kotoshi-wa</i> (exo1) _{NOM} isogashiku-te (exo1) _{NOM} (*2) _{ACC} mi-ni (*2) _{DAT} ike-masen-deshita. “But (I) _{NOM} couldn’t go (*2) _{DAT} to see (it=*2) _{ACC} this year because (I) _{NOM} was busy. ”

Figure 1: Chat Dialogue Example and Its Predicate-Argument Structure Annotation
Lower lines denote glosses of the upper lines. The bold words denote predicates, the square brackets [] denote intra-sentential arguments, and the round brackets () denote inter-sentential or exophoric arguments.

Corpus	Set	# of Articles /Dialogues	# of Sentences /Utterances	# of Words (per Sentence)	# of Predicates (per Sentence)
NAIST Text Corpus	Training	1,751	24,283	664,898 (27.4)	68,602 (2.83)
	Development	480	4,833	136,585 (28.3)	13,852 (2.87)
	Test	696	9,284	255,624 (27.5)	26,309 (2.83)
Chat Dialog Corpus	Training	184	6,960	61,872 (8.9)	7,470 (1.07)
	Test	101	4,056	38,099 (9.4)	5,333 (1.31)

Table 1: Sizes of Corpora

Case	Corpus	# of Arguments	Dep	Zero-Intra	Zero-Inter	Exophora		
						exo1	exo2	exog
Nominative	NAIST	68,598	54.5%	17.3%	11.4%	2.0%	0.0%	14.7%
	Dialogue	7,467	31.8%	7.4%	12.6%	23.9%	5.6%	18.8%
Accusative	NAIST	27,986	89.2%	6.9%	3.4%	0.0%	0.0%	0.4%
	Dialogue	1,901	46.6%	12.8%	27.5%	0.8%	0.1%	12.2%
Datative	NAIST	6,893	84.7%	10.2%	4.3%	0.0%	0.0%	0.8%
	Dialogue	2,089	37.6%	7.8%	15.0%	2.5%	1.1%	36.1%

Table 2: Distribution of Arguments in Training Corpora

Table 1 shows the statistics of the NAIST Text Corpus and the Chat Dialogue Corpus we created³. The size of the Dialogue Corpus is about 10% of the NAIST Corpus. The NAIST Corpus is divided into three parts: training, development, and test. The Dialogue Corpus is divided into training and test.

Table 2 shows distributions of arguments in the training sets of the NAIST/Dialogue corpora. We classified the arguments into the following six categories because each argument presents different difficulties for analysis by its position and syntactic relation. The first two categories (Dep and Zero-Intra) are the ones that in which the predicate and the argument occupy the same sentence.

- Dep: The argument directly depends on the predicate and vice versa on the parse tree.
- Zero-Intra: Intra-sentential zero-pronoun. The predicate and the argument are in the same sentence, but there is no direct dependency.
- Zero-Inter: Inter-sentential zero-pronoun. The predicate and the argument are in different sentences.
- exo1/exo2/exog: These are exophoric and denote zero-pronouns of the first person, second person, and the others (general), respectively.

By Table 2, we can see that the ratios of Dep in all cases decreased in the Dialogue Corpus. In the other categories, the tendencies between the nominative case and the accusative/dative cases were different. In the nominative case, the Zero-Intra also decreased in the Dialogue Corpus, and the declines were

³We regard a dialogue and an utterance as an article and a sentence, respectively.

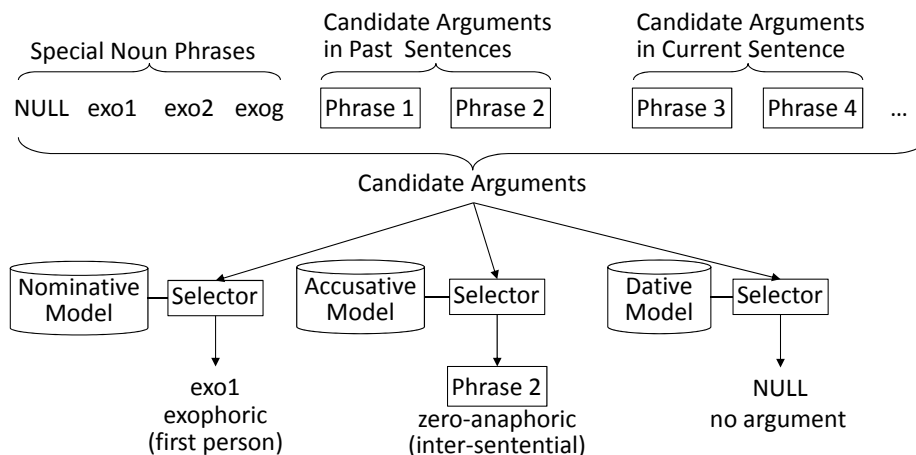


Figure 2: Structure of Argument Identification and Classification

assigned to `exo1` and `exo2`. Namely, the arguments in a sentence were reduced, and zero-pronouns increased compared with the newspaper articles. Note that many antecedents were the first or second person. On the other hand, in the accusative and dative cases, the declines of the `Dep` were assigned to the `Zero-Inter` or the `exog` in the Dialogue Corpus. Namely, anaphora resolution across multiple sentences is important to dialogue analysis. In contrast, most arguments and the predicate appear in the same sentence in the accusative/dative cases of newspapers.

4 Basic Strategy for Predicate-Argument Structure Analysis and Zero-Anaphora Resolution

4.1 Architecture

We use Imamura et al. (2009)'s method developed for newspaper articles as the base PASA in this paper. It can simultaneously identify arguments of a predicate in the sentence, those in other sentences, and exophoric arguments. The analyzer receives the entire article (dialogue) and performs the following steps for each sentence (utterance).

1. The input sentences are tagged and parsed. During parsing, the base phrases and their headwords are also identified. At this time, the part-of-speech tags and the parse trees of the Dialogue Corpus are supplied by applying the morphological analyzer MeCab (Kudo et al., 2004) and the dependency parser CaboCha (Kudo and Matsumoto, 2002). The NAIST Corpus version 1.5 already includes the part-of-speech tags and the parse trees.
2. Predicate phrases are identified from the sentences. We use the correct predicates in the corpora for the evaluation. When we build dialogue systems on PASA, predicate phrases will be identified using part-of-speech patterns that include verbs, adjectives, and copular verbs.
3. For each predicate, candidate arguments are acquired from the sentence that includes the predicate (called the current sentence) and the past sentences. Concretely, the following base phrases are regarded as candidates.
 - All noun phrases in the current sentence are extracted as intra-sentential candidates regardless of syntactic relations.
 - From the past sentences, noun phrases are contextually extracted as inter-sentential candidates. Details are described in Section 4.4.
 - Exophoric labels (`exo1`, `exo2`, and `exog`) and the `NULL` (the argument is not required) are added as special noun phrases.

4. The features are generated from the predicate phrase, candidate arguments, and their relations. The best candidate for each case is independently selected (Figure 2).

4.2 Models

The models for the selector are based on maximum entropy classification. The selector identifies the best noun phrase \hat{n} that satisfies the following equations from the candidate argument set \mathbf{N} .

$$\hat{n} = \operatorname{argmax}_{n_j \in \mathbf{N}} P(d(n_j) = 1 | X_j; M_c) \quad (1)$$

$$P(d(n_j) = 1 | X_j; M_c) = \frac{1}{Z_c(X)} \exp \sum_k \{ \lambda_{ck} f_k(d(n_j) = 1, X_j) \} \quad (2)$$

$$Z_c(X) = \sum_{n_j \in \mathbf{N}} \exp \sum_k \{ \lambda_{ck} f_k(d(n_j) = 1, X_j) \} \quad (3)$$

$$X_j = \langle n_j, v, A \rangle \quad (4)$$

where n denotes a candidate argument, \mathbf{N} denotes a set of candidate arguments of predicate v , $d(n)$ is a function that returns 1 iff candidate n becomes the argument, and M_c denotes the model of case c . In addition, $f_k(d(n_j) = 1, X_j)$ is a feature function, λ_{ck} denotes a weight parameter of the feature function, and A denotes the article from which all sentences are parsed.

Training phase optimizes the weight parameters in order to maximize the difference in posterior probabilities among the correct noun phrase and the other candidates. Specifically, the model of case M_c is learnt by minimizing the following loss function ℓ_c .

$$\ell_c = - \sum_i \log P(d(n_i) = 1 | X_i; M_c) + \frac{1}{2C} \sum_k \|\lambda_{ck}\|^2 \quad (5)$$

where n_i denotes the correct noun phrase of the i -th predicate in the training set, X_i denotes the i -th tuple of the correct noun phrase, the predicate, and the article $\langle n_i, v_i, A_i \rangle$. Since the posterior probability is normalized for each set of candidate arguments of a predicate by Equation (3), the probability of the correct noun phrase approaches closer to 1.0, and the probabilities of the other candidates approach closer to 0.0 in Equation (5).

4.3 Features

Similar to other studies (e.g., (Gildea and Jurafsky, 2002)), we use three types of features: 1) predicate features, 2) noun phrase (NP) features, and 3) the relationship between predicates and noun phrases (Table 3). We also introduce combined features of the ‘Noun’ with all other binary features because the features aim to select the best noun phrase.

The special features in this paper are the dependency language models (three types) and the obligatory case information (‘Frame’ feature), which are automatically acquired from large text corpora. We discuss them in Section 5.2.

4.4 Context Processing

Contexts of dialogues and newspaper articles are different. We should employ context processing specialized for the dialogues. However, contexts, including system and user utterances, should be managed collectively by the dialogue manager from the viewpoint of dialogue systems. Thus, this study uses the same context processing for the newspaper articles and dialogues. Note that the method in this paper controls the context by selecting the inter-sentential candidates. We can easily alter context management by providing candidate arguments from an external manager.

Context processing in this paper is as follows.

- From the current sentence, trace back to the past, and find a sentence that contains the other predicate (we call this the prior sentence). This process aims to ignore utterances that do not contain predicates.

Type	Name	Value	Remark
Predicate	Pred	Binary	Lemma of the predicate.
	PType	Binary	Type of predicate. One of ‘verb’, ‘adjective’, and ‘copular verb’.
	Voice	Binary	Declarative or not. If not, the passive/causative auxiliary verb is assigned.
	Suffix	Binary	Sequence of the functional words of the main clause. This feature aims to reflect the speech act of the utterance.
	Frame	Binary	Obligatory case information. The case requires argument (1) or not (0).
Noun Phrase	Noun	Binary	Headword of the NP
	Particle	Binary	Case particle of the base phrase. If the NP is a special noun phrase, this is NULL.
	NType	Binary	If the substance of the NP is in the article, this is ‘NP’; otherwise the same value of the ‘Noun’ feature.
	Surround	Binary	POS tags of the surrounding words of the NP. The window size is ± 2 .
Relation between Predicate and NP	PhPosit	Binary	Distance between the predicate and the NP. If they are in different sentences, or the NP is an exophora, this is NULL.
	Syn	Binary	Dependency path between the predicate and the NP. If they are in different sentences, or the NP is an exophora, this is NULL.
	Speaker	Binary	Whether the speakers of the predicate and the NP are the same (SAME) or not (OTHER).
Dependency Language Models	$\log P(n c, v)$	Real	Generation probability of NP n given predicate v and case c .
	$\log P(v c, n)$	Real	Generation probability of predicate v given NP n and case c .
	$\log P(c n)$	Real	Generation probability of case c given NP n .

Table 3: List of Features

- All noun phrases that lie between the prior to the current sentence are added to the candidate arguments. In addition, noun phrases that are used as arguments of any predicates are also added (called argument recycling (Imamura et al., 2009)). Argument recycling covers wide contexts because it can employ distant noun phrases if the past predicates have inter-sentential arguments.

5 Adaptation to Chat Dialogues

The method described in the previous section is common to dialogues and newspaper articles. This section describes the adaptation made to target dialogues.

5.1 Adaptation of Model Parameters

In order to tune the difference in the argument distribution, model parameters of the selectors are adapted to the dialogue domain. We use the feature augmentation method (Daumé, 2007) as the domain adaptation technique; it has the same effect as regarding the source domain to be prior knowledge, and the parameters are optimized to the target domain. Concretely, the models of the selectors are learnt and applied as follows.

1. First, the feature space is segmented into three parts: common, source, and target.
2. The NAIST Corpus and the Dialogue Corpus are regarded as the source and the target domains, respectively. The features from the NAIST Corpus are deployed to the common and the source spaces, and those from the Dialogue Corpus are deployed to the common and the target spaces.
3. The parameters are estimated in the usual way on the above feature space. The weights of the common features are emphasized if the features are consistent between the source and target. With regard to domain-dependent features, the weights in the respective space, source or target, are emphasized.
4. When the argument is identified, the selectors use only the features in the common and target spaces. The parameters in the spaces are optimized to the target domain, plus we can utilize the features that appear only in the source domain data.

5.2 Weak Knowledge Acquisition from Very Large Resources

In this paper, we use two types of knowledge to reduce the harmful effect of out-of-vocabulary in the training corpus. Both types are constructed by automatically analyzing, summing up, and filtering large

text corpora (Kawahara and Kurohashi, 2002; Sasano et al., 2008; Sasano et al., 2013). They provide information about unknown words with some confidence but they do contain some errors. We use them as the features of the models, and parameters are optimized by the discriminative learning of the selectors.

5.2.1 Obligatory Case Information (Frame Feature)

Case frames are important clues for SRL and PASA. The obligatory case information (OCI) comprises subsets of the case frames that only clarify whether the cases of each predicate are necessary or not.

The OCI dictionary is automatically constructed from large text corpora as follows. The process assumes that 1) most of the cases match the case markers if the noun phrase directly depends on the predicate, and 2) if the case is obligatory, the occurrence rate on a specific predicate is higher than the average rate of all predicates.

1. Similar to PASA in this paper (c.f., Section 4.1), predicates and base phrases are identified by tagging and parsing raw texts.
2. Noun phrases that directly depend on the predicate and accompany a case marker are extracted. We sum up the frequency of the predicate and cases.
3. Highly frequent predicates are selected according to the final dictionary size. Obligation of the cases is determined so as to satisfy the following two conditions.
 - Co-occurrence of the predicate and the case $\langle v, c \rangle$ are higher than the significance level ($p \leq 0.001$; $LLR \geq 10.83$) by the log-likelihood-ratio test.
 - The case of the predicate appears at least 10% more frequently than the average of all predicates.

We constructed two OCI dictionaries. The Blog dictionary contains about 480k predicates from one year of blogs (about 2.3G sentences,). The News dictionary contains about 200k predicates from 12 years of newspaper articles (about 7.7M sentences). The coverage of predicates in the training set of the Dialogue Corpus was 98.5% by the Blog dictionary and 96.4% by the News dictionary.

5.2.2 Dependency Language Models

Dependency language models (LMs) represent semantic/pragmatic collocations among predicate v , case c , and noun phrase n . The generation probabilities of v , c , and n are computed by n -gram models. More concretely, the following real values are computed. The purpose of the biases (probabilities involved $\langle \text{unk} \rangle$) is to correct the values to be positive.

- $\log P(n|c, v) - \log P(\langle \text{unk} \rangle|c, v)$
- $\log P(v|c, n) - \log P(v|c, \langle \text{unk} \rangle)$
- $\log P(c|n) - \log P(c|\langle \text{unk} \rangle)$

Each dependency LM is constructed from the tuples of $\langle v, c, n \rangle$ extracted in Section 5.2.1 using the SRILM (Stolcke et al., 2011). Note that since the obligatory case information corresponds to the generation probability of the case ($P(c|v)$), we exclude it from the dependency LMs.

Similar to the OCI dictionaries, we constructed two sets of dependency language models from the Blog and the News sentences. The coverage of triples $\langle v, c, n \rangle$ appeared in the training set of the Dialogue Corpus was 76.4% by the Blog LMs and 38.3% by the News LMs. The Blog LMs cover the Dialogue Corpus more comprehensively than the News LMs.

6 Experiments

We evaluate the accuracies of the proposed PASA on the Dialogue Corpus (Table 1) from the perspectives of parameter adaptation and the effect of the automatically acquired knowledge. The evaluation metric is F-measure of each case (includes exophora identification).

Case	Type	# of Args.	a) Adaptation OCI:Blog LMs:Blog	b) NAIST♡ OCI:Blog LMs:Blog	c) Dialogue◇ OCI:Blog LMs:Blog	d) Adaptation OCI:News♠ LMs:Blog	e) Adaptation OCI:Blog LMs:News♣
Nominative	Dep	1,811	83.3% ♡◇	77.6%	82.7%	83.0%	82.7%
	Zero-Intra	511	37.4%	43.7% ♡	36.6%	36.5%	38.1%
	Zero-Inter	767	8.6%♣	9.1%	9.0%	8.3%	4.5%
	exo1	1,193	70.2%♡	13.5%	69.9%	70.1%	70.3%
	exo2	281	46.8%♡◇	0.0%	43.1%	47.2%	46.8%
	exog	767	46.8%♡	32.5%	27.9%	47.2%	47.7% ♣
Total	5,330	61.5% ♡	44.4%	61.1%	61.4%	61.4%	
Accusative	Dep	614	84.2% ♡◇♣	78.6%	81.5%	84.2%	82.4%
	Zero-Intra	149	42.9%♡♠♣	27.1%	45.0%	38.9%	34.3%
	Zero-Inter	399	30.4%♡♣	0.5%	30.9%	29.4%	24.3%
	exo1	19	0.0%	0.0%	0.0%	9.5%	10.0%
	exo2	7	0.0%	0.0%	0.0%	0.0%	0.0%
	exog	98	25.6%♡	0.0%	27.9%	25.2%	25.6%
Total	1,286	59.0% ♡♣	51.6%	58.9%	58.4%	56.0%	
Dative	Dep	566	80.5%♡◇	54.0%	79.0%	80.1%	80.7%
	Zero-Intra	70	20.7% ♡♣	0.0%	20.0%	20.7%	11.8%
	Zero-Inter	169	14.6%♡	0.0%	14.8%	14.4%	13.4%
	exo1	32	0.0%	0.0%	0.0%	0.0%	0.0%
	exo2	4	0.0%	0.0%	0.0%	0.0%	0.0%
	exog	265	45.4% ♡◇	0.0%	43.1%	44.0%	44.9%
Total	1,106	58.6% ♡◇	32.2%	57.2%	58.2%	58.4%	

Table 4: F-measures among Methods/OCI dictionary/Dependency LMs on Dialogue Test Set. The bold values denote the highest F-measures among all methods. The marks ♡, ◇, ♠, ♣ denote significantly better methods by comparing a) with b), c), d), and e), respectively. We used the bootstrap resampling method (1,000 iterations) as the significance test, in which the significance level was 0.05.

6.1 Experiment 1: Effect of Parameter Adaptation

We compared three methods in order to evaluate parameter adaptation: a) The feature augmentation is applied to the training (Adaptation). b) Only the NAIST Corpus is used for training (NAIST Training). c) Only the Dialogue Corpus is used (Dialogue Training). The NAIST Training corresponds to a conventional PASA for newspaper articles. The results on the Dialogue test set are shown in the 4th, 5th, and 6th columns in Table 4.

First, comparing methods a) Adaptation and b) NAIST training, Adaptation was better than the NAIST training for most types (The ♡ mark denotes ‘significantly better’). In particular, the total F-measures of all cases were significantly better than NAIST training. Focusing on the types of arguments, the most characteristic results were exophoras of the first/second persons (exo1 and exo2) of the nominative case. These two types dominate of the nominative case (about 28%), and exo1 (70.2%) and exo2 (46.8%) became analyzable. Other types such as the Zero-Inter and the exog of the accusative and dative cases, which could not be analyzed by NAIST training, became analyzable.

Comparing methods a) Adaptation and c) Dialogue training (c.f., ◇), the F-measures of Dialogue training approached those of Adaptation even though the size of the Dialogue Corpus was small. Only the F-measure of the dative case of Adaptation was significantly better than Dialogue training in total. This does not imply that the corpus size is sufficient. Rather, we suppose that the Adaptation strategy could not adequately utilize the advantages of the NAIST Corpus. Adding more dialogue data would further improve the accuracies on the Dialogue test set.

6.2 Experiment 2: Differences among Automatically Acquired Knowledge

The columns a), d), and e) in Table 4 show the results for the proposed method (Adaptation). Note that the combination of the OCI dictionary and the dependency language models were changed to a) (Blog, Blog), d) (News, Blog), and e) (Blog, News).

When the OCI dictionary was changed from a) Blog to d) News (c.f., ♠), there were no significant differences in almost all types except for the Zero-Intra of the accusative case. We suppose that this

is because the coverage of the Blog and News dictionaries were almost the same, and obligatory cases of predicates are general information regardless of the domain.

On the contrary, when the dependency LMs were changed from a) Blog to e) News (c.f., ♣), the F-measures of some types significantly dropped, especially the Zero-Intra and Zero-Inter types, which are strongly influenced by semantic relation. For example, the Zero-Inter type of the accusative case was changed from 30.4% to 24.3%, and the F-measure consequently decreased by 3.0 points in total in the accusative case. Zero-anaphora resolution cannot rely on syntax, and the dependency LMs that measure semantic collocation become relatively important. The Blog LMs yielded greater coverage than the News LMs in this experiment. We can conclude that high-coverage LMs are better for improving the zero-anaphora resolution.

7 Conclusion

This paper presented predicate-argument structure analysis with zero-anaphora resolution for dialogues. We regarded this task as a kind of domain adaptation from newspaper articles, which are conventionally studied, to dialogues. The model parameters were adapted to the dialogues by using a domain adaptation technique. In order to address the out-of-vocabulary issue, the obligatory case information and the dependency language models were constructed from large text corpora and applied to the selectors.

As a result, arguments that could not be analyzed by PASA for newspaper articles (e.g., zero-pronouns of the first and second persons in the nominative case) became analyzable by adding only a small number of dialogues. The parameter adaptation achieved some improvement. Moreover, we confirmed that high-coverage dependency LMs contribute to improving zero-anaphora resolution and the overall accuracy.

Although we focused on parameter distribution and out-of-vocabulary in this paper, there are the other differences between dialogues and newspaper articles. For example, we did not discuss the exchange of turns, which is a special phenomenon of dialogues. To consider further phenomena is our future work. We are also evaluating the effectiveness of our PASA by incorporating it into a dialogue system (Higashinaka et al., 2014).

References

- Xavier Carreras and Lluís Màrquez. 2004. Introduction to the CoNLL-2004 shared task: Semantic role labeling. In Hwee Tou Ng and Ellen Riloff, editors, *HLT-NAACL 2004 Workshop: Eighth Conference on Computational Natural Language Learning (CoNLL-2004)*, pages 89–97, Boston, Massachusetts, USA, May.
- Xavier Carreras and Lluís Màrquez. 2005. Introduction to the CoNLL-2005 shared task: Semantic role labeling. In *Proceedings of the Ninth Conference on Computational Natural Language Learning (CoNLL-2005)*, pages 152–164, Ann Arbor, Michigan, June.
- Hal Daumé, III. 2007. Frustratingly easy domain adaptation. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 256–263, Prague, Czech Republic, June.
- Matthew Gerber and Joyce Y. Chai. 2012. Semantic role labeling of implicit arguments for nominal predicates. *Computational Linguistics*, 38(4):755–798.
- Daniel Gildea and Daniel Jurafsky. 2002. Automatic labeling of semantic roles. *Computational Linguistics*, 28(3):245–288.
- Ryuichiro Higashinaka, Kenji Imamura, Toyomi Meguro, Chiaki Miyazaki, Nozomi Kobayashi, Hiroaki Sugiyama, Toru Hirano, Toshiro Makino, and Yoshihiro Matsuo. 2014. Towards an open domain conversational system fully based on natural language processing. In *Proceedings of the 25th International Conference on Computational Linguistics (COLING 2014)*, Dublin, Ireland, August.
- Eduard Hovy, Mitchell Marcus, Martha Palmer, Lance Ramshaw, and Ralph Weischedel. 2006. OntoNotes: The 90% solution. In *Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers*, pages 57–60, New York City, USA, June.
- 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, Prague, Czech Republic, June.

- Kenji Imamura, Kuniko Saito, and Tomoko Izumi. 2009. Discriminative approach to predicate-argument structure analysis with zero-anaphora resolution. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, pages 85–88, Singapore, August.
- Zheng Ping Jiang and Hwee Tou Ng. 2006. Semantic role labeling of NomBank: A maximum entropy approach. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pages 138–145, Sydney, Australia, July.
- Daisuke Kawahara and Sadao Kurohashi. 2002. Fertilization of case frame dictionary for robust Japanese case analysis. In *Proceedings of the 19th International Conference on Computational Linguistics (COLING-2002)*, pages 425–431, Taipei, Taiwan, August.
- Mamoru Komachi, Ryu Iida, Kentaro Inui, and Yuji Matsumoto. 2007. Learning-based argument structure analysis of event-nouns in Japanese. In *Proceedings of the Conference of the Pacific Association for Computational Linguistics (PACLING)*, pages 208–215, Melbourne, Australia, September.
- Taku Kudo and Yuji Matsumoto. 2002. Japanese dependency analysis using cascaded chunking. In *CoNLL 2002: Proceedings of the 6th Conference on Natural Language Learning 2002 (COLING 2002 Post-Conference Workshops)*, pages 63–69, Taipei, Taiwan.
- Taku Kudo, Kaoru Yamamoto, and Yuji Matsumoto. 2004. Applying conditional random fields to Japanese morphological analysis. In Dekang Lin and Dekai Wu, editors, *Proceedings of EMNLP 2004*, pages 230–237, Barcelona, Spain, July.
- Egoitz Laparra and German Rigau. 2013. ImpAr: A deterministic algorithm for implicit semantic role labelling. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1180–1189, Sofia, Bulgaria, August.
- Lluís Màrquez, Xavier Carreras, Kenneth C. Litkowski, and Suzanne Stevenson. 2008. Semantic role labeling: An introduction to the special issue. *Computational Linguistics*, 34(2):145–159.
- Yuichiro Matsubayashi, Ryu Iida, Ryohei Sasano, Hikaru Yokono, Suguru Matsuyoshi, Atsushi Fujita, Yusuke Miyao, and Kentaro Inui. 2014. Issues on annotation guidelines for Japanese predicate-argument structures. *Journal of Natural Language Processing*, 21(2):333–377, April. in Japanese.
- Martha Palmer, Daniel Gildia, and Paul Kingsbury. 2005. The proposition bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31(1):71–105.
- Sameer S. Pradhan, Wayne Ward, and James H. Martin. 2008. Towards robust semantic role labeling. *Computational Linguistics*, 34(2):289–310.
- Sameer Pradhan, Alessandro Moschitti, and Nianwen Xue, editors. 2012. *Joint Conference on EMNLP and CoNLL: Proceeding of the Shared Task: Modeling Multilingual Unrestricted Coreference in Onto Notes*, Jeju, Korea, July.
- Ryohei Sasano, Daisuke Kawahara, and Sadao Kurohashi. 2008. A fully-lexicalized probabilistic model for Japanese zero anaphora resolution. In *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*, pages 769–776, Manchester, UK, August.
- Ryohei Sasano, Daisuke Kawahara, Sadao Kurohashi, and Manabu Okumura. 2013. Automatic knowledge acquisition for case alternation between the passive and active voices in Japanese. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1213–1223, Seattle, Washington, USA, October.
- Andreas Stolcke, Jing Zheng, Wen Wang, and Victor Abrash. 2011. SRILM at sixteen: Update and outlook. In *Proceedings of IEEE Automatic Speech Recognition and Understanding Workshop (ASRU 2011)*, Waikoloa, Hawaii, December.
- Hirotoishi Taira, Sanae Fujita, and Masaaki Nagata. 2008. A Japanese predicate argument structure analysis using decision lists. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pages 523–532, Honolulu, Hawaii, October.
- Katsumasa Yoshikawa, Masayuki Asahara, and Yuji Matsumoto. 2011. Jointly extracting Japanese predicate-argument relation with markov logic. In *Proceedings of 5th International Joint Conference on Natural Language Processing*, pages 1125–1133, Chiang Mai, Thailand, November.