

Effectiveness of Domain Adaptation in Japanese Predicate-Argument Structure Analysis

Mizuki Sango, Hitoshi Nishikawa, Takenobu Tokunaga

Tokyo Institute of Technology / W8-73, 2-12-1 Ookayama, Meguro, Tokyo 152-8552, Japan
 {sango.m.ab@m, hitoshi@c, take@c}.titech.ac.jp

Abstract

This paper proposes introducing domain adaptation into Japanese predicate-argument structure (PAS) analysis. Our investigation of a Japanese balanced-corpus revealed that the distribution of argument types is different across text media, particularly the difference is significant when the argument is exophoric. The past Japanese PAS analysis research has disregarded this tendency. We start with an RNN-based PAS analyzer as a baseline, extending it by introducing three kinds of domain-adaptation techniques and their combinations. The evaluation experiments using a Japanese balanced-corpus (BCCWJ) confirmed that the domain-adaptation is effective for improving the performance of the Japanese PAS analysis.

1 Introduction

Predicate-argument structure (PAS) analysis is the task to identify the argument for each case of the target predicate. As it is a fundamental analysis for various natural language processing (NLP) applications, the PAS analysis has been one of the most active research areas in NLP. In discourse-oriented languages like Japanese, the target language of the present study, arguments are often omitted from the sentence Kayama (2003). Those omitted arguments are considered as zero-pronouns or exophora.

- (1) *meiru-o kaite okuttayo. yondene.*
 mail_{ACC} wrote_{v₁} sent_{v₂} read_{v₃}/imperative
 I wrote a mail to you and sent it to you. Read it.

For instance, example (1) has three predicates (v_1 , v_2 and v_3) and one explicit argument (mail).

predicate \ case	NOM	ACC	DAT
v_1 : wrote	[writer]	mail	[reader]
v_2 : sent	[writer]	(mail)	[reader]
v_3 : read	[reader]	((mail))	none

Table 1: PAS analysis result for example (1)

The PAS analysis result of example (1) looks like Table 1, where the elements enclosed with square brackets are exophoric, the round bracketed is an intra-sentential zero-anaphora and the double round bracketed is an inter-sentential zero-anaphora. The accusative argument of v_1 , “*meiru-o* (mail)”, is explicitly marked by the case marker “*o*” and has a dependency relation with v_1 , which is indicated by a bare noun, i.e. without any bracket.

Although the Japanese PAS analysis is similar to the semantic role labeling (SRL) (Zhou and Xu, 2015; He et al., 2017), it also involves anaphora resolution for zero-pronouns and exophora to identify the argument for every case of the predicate, which corresponds to filling the bracketed elements in Table 1. We also find omitted arguments in other pro-drop languages such as Chinese, Turkish, and some null-subject languages in the Romance languages (Iida and Poesio, 2011; Rello et al., 2012; Chen and Ng, 2016; Yin et al., 2017).

The past Japanese PAS analysis utilizes various features obtained from the morphological and syntactic analysis (Matsubayashi and Inui, 2017; Hayashibe et al., 2011; Imamura et al., 2014; Shibata et al., 2016; Ouchi et al., 2015; Yoshikawa et al., 2013; Taira et al., 2008). The recent approach includes the end-to-end approach that does not require any intermediate analysis (Ouchi et al., 2017).

The contribution of the present paper to the Japanese PAS analysis is twofold. Firstly we subcategorize the exophora into fine-grained classes, namely, the exophoric text writer (exo1), reader (exo2) and the other entity (exoX). Example (2) depicts the necessity of the subcategorization.

- (2) *sandoitti taberu.*
 sandwich eat
 I eat sandwich. / Do you eat sandwich?

Both the exophoric speaker (exo1) and hearer (exo2) can be the nominative argument of the verb “eat” and accordingly the sentence meaning is different. To distinguish these two meanings, the subcategorization of the exophora is necessary.

Secondly, we introduce domain-adaptation techniques into the Japanese PAS analysis. Surdeanu et al. (2008) and Hajič et al. (2009) reported that the SRL performance degraded when the domains were different between the training and testing data. Yang et al. (2015) tackled this problem by introducing the domain adaptation into a deep learning method. As most of the past studies of the Japanese PAS analysis targeted a mono-type of texts, i.e. newspaper articles, the domain adaptation did not matter, except for Imamura et al. (2014). They trained the PAS analyzer for dialogues by using newspaper articles. However, pairs of other media types have not been investigated yet. In contrast, we target various types of Japanese texts; we use Balanced Corpus of Contemporary Written Japanese (BCCWJ)¹ (Maekawa et al., 2014) for evaluation. BCCWJ contains 100 million words that were systematically collected from several source media such as newspaper articles (PN), books (PB), magazines (PM), white papers (OW), QA texts in the Internet (OC) and blog texts (OY). We use the core data set of BCCWJ consisting of about two million words annotated with co-reference and predicate-argument relations for nominative, dative and accusative cases. As we describe in the next section, the distribution of exophoric arguments is notably different over the source media; thus consideration of the difference in the source media is necessary.

We start with a recurrent neural network (RNN)-based base model and extend it by introducing the

¹http://pj.ninjal.ac.jp/corpus_center/bccwj/en/

following five kinds of domain adaptation. (1) The fine-tuning method trains the model with the entire training data and uses the learnt parameters as the initial parameter values for the second stage learning with the target-domain training data. (2) The feature augmentation method trains a shared network and domain-specific networks simultaneously (Kim et al., 2016). (3) The class probability shift method skews the output probability of the network based on the prior probability distribution of the argument types across the domains. (4) The voting method determines the output by the majority of the above three methods. (5) The mixture method combines the fine-tuning method, the feature augmentation method and the class probability shift method into a single method. We describe the details of each method in section 4.

2 Problem setting

2.1 Argument type

The past Japanese PAS analysis targeted various combinations of argument types. Table 2 summarizes the previous studies and their target argument types. The table header represents the classification of arguments from a linguistic viewpoint. Arguments are divided into endophora and exophora depending on whether they appear in the text or not. The endophoric arguments are further divided into intra- and inter-sentential arguments depending on whether they appear in the same sentence as the predicate. Some intra-sentential arguments have a dependency relation with the predicate, but this is not always the case. We call the latter case intra-sentential zero-anaphora. Since the inter-sentential arguments do not have a dependency relation with the predicate, they are also zero-anaphoric. As we described in the previous section, we divide the exophoric arguments into three subcategories: writer, reader and the other entity. In what follows, we use the labels shown in Table 2 for denoting argument types. The label none indicates that the predicate takes no argument for that case. For instance, intransitive verbs do not take an accusative argument; thus the accusative case of intransitive verbs should be filled with none.

Table 2 shows that the inter-sentential arguments were tackled by the fewer studies than the intra-

work	\	label	endophora			exophora			
			intra-sentential		inter-sentential	exo1	exo2	exoX	none
			non-zero	zero	inter				
			intra(dep)	intra(zero)					
present work			○	○		○	○	△	○
Matsubayashi and Inui (2017)			○	○					
Ouchi et al. (2017)			○	○					
Shibata et al. (2016)			○	○		○	○		○
Imamura et al. (2014)			○	○	○	○	○	○	○
Hangyo et al. (2013)				○	○	○	○	○	
Yoshikawa et al. (2013)			○	○					
Hayashibe et al. (2011)			○	○	○				○
Sasano and Kurohashi (2011)				○	○			○	○
Imamura et al. (2009)			○	○	○	△	△	△	△

Table 2: Target argument types of the past studies

sentential arguments. Identifying the inter-sentential arguments requires searching in larger space compared with the intra-sentential arguments, and thus the problem becomes more difficult. Unlike the inter-sentential arguments, identifying the exophoric arguments, in particular, the *exo1* and *exo2* arguments do not drastically increase the search space. They are easy to be introduced into the PAS analyzer (Shibata et al., 2016; Hangyo et al., 2013). There are, however, variations in the *exoX* argument treatment. Hangyo et al. (2013) and Imamura et al. (2014) assume a single category for *exoX*, while Sasano and Kurohashi (2011) identifies a named-entity class of the exophoric entity. Imamura et al. (2009) does not distinguish the subcategories of the exophoric arguments and the *none* argument. This fact is indicated by \triangle in Table 2. In this present study, we target the intra-sentential arguments together with the exophoric argument. We do not distinguish the *exoX* and *inter* arguments; they are treated as a single category unknown. This is the reason for \triangle at the *exoX* column in our work. As the target predicates, we use those which are marked as “predicate” with arguments in BCCWJ and event nouns which usually become verbs when being used with light verbs.

2.2 Domain dependency of argument type distribution

The previous studies on the Japanese PAS analysis dealt with the texts from a single “domain” in

a broad sense; many of them used newspaper articles. We use BCCWJ for the evaluation of the proposed methods, considering six source media in BCCWJ (OW, PB, PM, OW, OC, OY) as independent domains. It is probable that the characteristics of texts are different across the source media and therefore the PAS analysis performance might be affected by the domain characteristics. One of our objective in this study is to confirm that introducing domain-adaptation is effective for the Japanese PAS analysis.

The length of sentences is different across the source media. The sentence length affects the distance between the predicate and its arguments. Table 3 shows the distribution of the argument type for each case across the six source media. The OW texts have fewer inter-sentential arguments than intra-sentential arguments in contrast with the other media. We can partially explain this difference by the average sentence length. A longer sentence has more chance to include the arguments of the predicate within the sentence. The distribution of the exophoric arguments (the shaded rows) for the nominative case, in particular, that of *exo1* and *exo2* is notably different across the media. Both *exo1* and *exo2* in OC show quite high numbers compared with the other domain. The OC contains QA texts that are similar to dialogues. It is natural not to explicitly mention a questioner (*exo1*) and responders (*exo2*). The OY texts also show high numbers at *exo1* and *exo2* but unlike OC they are skewed toward *exo1*, as the blog texts are monologues.

case	arg. type \ media # of predicates	OC 16,824	OY 15,612	OW 33,529	PB 32,532	PM 30,410	PN 47,609	All 176,516
NOM	none	0.06	0.74	0.19	0.39	0.76	1.00	0.58
	intra(dep)	37.33	35.86	35.75	46.20	43.31	43.06	41.11
	intra(zero)	10.50	13.36	18.20	18.03	14.62	19.75	16.81
	inter	18.25	13.21	8.34	18.46	21.68	18.35	16.58
	exo1	12.13	19.16	0.26	0.69	1.60	0.69	3.49
	exo2	8.26	2.59	0.03	0.32	1.33	0.57	1.46
	exoX	13.46	15.05	37.22	15.88	16.61	16.54	19.93
ACC	none	62.73	69.59	45.91	61.29	62.74	59.95	59.13
	intra(dep)	21.43	21.27	41.10	28.41	28.70	31.36	30.37
	intra(zero)	4.57	3.55	5.73	4.66	3.95	3.97	4.45
	inter	7.06	3.59	2.24	3.48	3.07	3.14	3.43
	exo1	0.17	0.25	0.00	0.00	0.03	0.01	0.05
	exo2	0.10	0.02	0.00	0.01	0.01	0.00	0.02
	exoX	3.88	1.68	4.96	2.12	1.41	1.44	2.48
DAT	none	80.12	87.37	87.12	81.43	85.05	85.71	84.69
	intra(dep)	10.78	9.08	9.03	13.43	12.42	11.23	11.19
	intra(zero)	1.97	1.15	1.65	1.90	1.20	1.44	1.55
	inter	3.14	1.21	0.75	2.41	1.07	1.16	1.49
	exo1	1.42	0.36	0.00	0.02	0.04	0.01	0.18
	exo2	0.75	0.17	0.00	0.02	0.03	0.03	0.11
	exoX	1.81	0.67	1.42	0.79	0.17	0.37	0.78

OC: QA texts, OY: blog texts, OW: white papers, PB: books, PM: magazines, PN: newspapers

Table 3: BCCWJ distribution of argument type across source media [%]

3 Deep recurrent model

We implement our Japanese PAS analysis method using a recurrent neural network (RNN) model consisting of the following three layers.

Input layer maps each word into a feature vector.

Hidden layer is a bi-directional RNN.

Output layer has a linear function and a softmax function for binary classification.

Since our model outputs a binary label for each word indicating whether the word is the argument of a case for the target predicate, we have to prepare a model for each case. As our preliminary experiment showed that a model solving three cases simultaneously was inferior to the models for each case, we adopt the individual model for each case in this experiment. We show our model in Figure 1, which is formalized as follow:

$$\bar{x} = w_a \oplus w_f \oplus b_f \quad (1)$$

$$h^1 = \text{BiLSTM}(\bar{x}) \quad (2)$$

$$h^2 = \text{linear}(h^1) \quad (3)$$

$$p = \text{softmax}(h^2) \quad (4)$$

Our model first receives a sequence of words as an input sentence. Sequence of words $\{w_t\}_0^T$ in

the input sentence is mapped into corresponding sequence of feature vectors $\{\bar{x}_t\}_0^T$. Feature vector \bar{x} is made by concatenating word embedding w_a , part-of-speech (POS) embedding w_f and syntactic features b_f . The feature vector \bar{x} is fed into a bi-directional long short-term memory recurrent neural network (BiLSTM) (Schuster and Paliwal, 1997; Graves et al., 2005). Then $\text{BiLSTM}(\cdot)$ computes and outputs vector h^1 for each word. Function $\text{linear}(\cdot)$ takes h^1 and outputs $h^2 = (h_0^2, h_1^2)$. Finally, Function $\text{softmax}(\cdot)$ takes h^2 and output probability p .

3.1 Input layer

We define three types of features: word embeddings, POS embeddings, and syntactic features.

Word embedding We use word embeddings developed from Japanese Wikipedia by Suzuki et al. (2016)²

POS embedding Each word has a hierarchical POS tag with at most six levels. We assign a five-dimensional random vector to each level of a POS tag. Thus a hierarchical POS tag is represented by

²Japanese Wikipedia Entity Vector http://www.cl.ecei.tohoku.ac.jp/~m-suzuki/jawiki_vector/

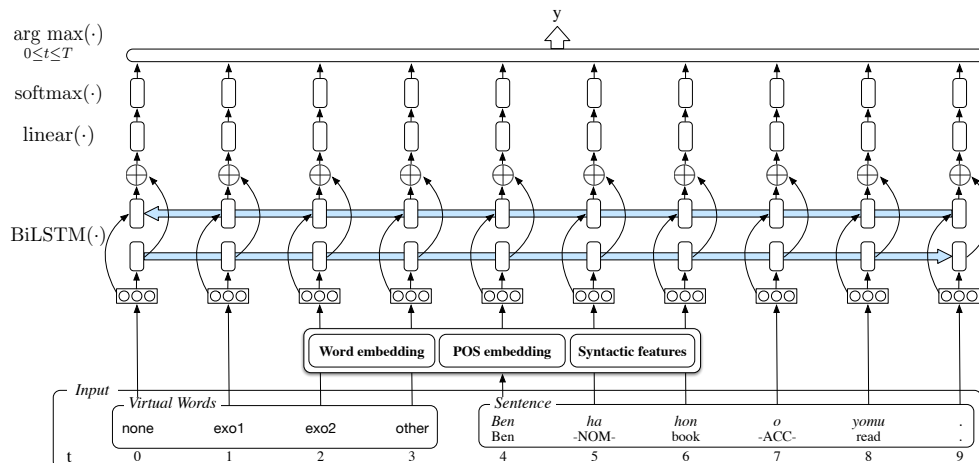


Figure 1: Deep recurrent model for Japanese PAS analysis

a 30-dimensional vector that is made by concatenating six vectors of each level. A POS tag vector with less than six levels is padded by zero-vectors.

Syntactic features Syntactic features include four kinds of features. (1) **Head** feature is a binary feature which indicates whether a word is the head of a base phrase or not. (2) **Position in the sentence** is an integer feature that indicates the phrase-based distance from the beginning of a sentence. We use the phrases annotated in the corpus. A word in the first phrase in the input sentence has value zero for this feature. (3) **Distance from the predicate** is an integer feature which indicates the distance from the target predicate to be analyzed. (4) **Target verb** is a binary feature which indicates whether a word is the predicate to be analyzed.

In order to allow our model to output the labels for none, exo1, exo2, and unknown, we add virtual words representing them before the first word in a sentence. We assign the feature for these virtual words as follows.

none We set a zero vector for none.

exo1 We use the word embedding of “*boku* (I)”, commonly-used first-person singular pronoun in Japanese for exo1.

exo2 We use the word embedding of “*omae* (you)” for exo2, which is commonly-used second-person singular pronoun in Japanese.

unknown We use the word embedding of “*kore* (this)” for unknown, which is commonly-used

third-person singular pronoun in Japanese.

3.2 Hidden layer

In the hidden layer, the forward LSTM ($LSTM^f$) computes h_t^f by the feature vector \bar{x}_t and its state h_{t-1}^f for each t . Conversely, the backward LSTM ($LSTM^b$) computes h_t^b by the feature vector \bar{x}_t and h_{t+1}^b for each t . BiLSTM concatenates h_t^f and h_t^b and outputs h_t^1 for each t .

$$\begin{aligned} h_t^1 &= \text{BiLSTM}(\bar{x}_t) \\ &= LSTM^f(\bar{x}_t, h_{t-1}^f) \\ &\oplus LSTM^b(\bar{x}_t, h_{t+1}^b) \end{aligned} \quad (5)$$

We then feed h_t^1 into function $\text{linear}(\cdot)$ to obtain two-dimensional vector h_t^2 .

$$h_t^2 = \text{linear}(h_t^1) \quad (6)$$

3.3 Output layer

In the output layer, our model judges whether the word is the case argument of the target predicate. Function $\text{softmax}(\cdot)$ translates two-dimensional vector h_t^2 into a probability indicating to what degree the word can be the case argument of the target predicate.

$$p_t = \text{softmax}(h_t^2) \quad (7)$$

p_t is a probability that t -th word is the case argument. Our model selects the word which has the highest probability p_y .

$$y = \arg \max_{0 \leq t \leq T} (p_t) \quad (8)$$

4 Domain adaptation

First, we prepare the following five baseline models. (1) Model `Each-D` is trained with the data of each media source. (2) Model `All` is trained with the entire data of all media sources. (3) Model `Small` is trained with the data reduced to an amount of 75% of all media sources, which is prepared to see the impact of the data size on accuracy. (4) Model `Out-D` is trained with the out-domain data. (5) Model `One-H` is trained with the entire. We extend each training example by adding a one-hot vector to indicate its media source. This model is a baseline for domain adaptation.

On top of these baseline models, we prepare five domain adaptation methods as follows.

(1) Fine-tuning We train the model with the entire data of all media sources to build the `All` model. We then train that model with the data of the target media source.

(2) Feature augmentation The second method follows Kim et al. (2016) in which BiLSTM^m is prepared for each media source m in addition to a common BiLSTM^c covering all media sources. This method is summarized as follows.

$$\bar{x} = w_a \oplus w_f \oplus b_f \quad (9)$$

$$h^1 = \text{BiLSTM}^m(\bar{x}) \oplus \text{BiLSTM}^c(\bar{x}) \quad (10)$$

$$h^2 = \text{linear}^m(h^1) \quad (11)$$

$$p = \text{softmax}(h^2) \quad (12)$$

We expect that the media-specific BiLSTM^m learns media-specific characteristics and the common BiLSTM^c learns general characteristics of the PAS analysis. We train this model with randomly selected batches from all media sources.

(3) Class probability shift The third method leverages the distribution of argument types for each case in each media source. Given a target case, we use probability p_{tp}^m that argument type tp appears in the training examples of media source m for the target case. Since this probability distribution is different among media sources, we use the difference in this distribution as follows. We define two functions

$f^m(h)$ and $g^m(h)$ for each media source m .

$$f^m(h) = \frac{p_{tp}^m}{p_{tp}^{\text{All}}} \cdot h \quad (13)$$

$$g^m(h) = \frac{100 - p_{tp}^m}{100 - p_{tp}^{\text{All}}} \cdot h, \quad (14)$$

Where tp is one of `none`, `exo1`, `exo2`, `unknown` and `intra`. Label `intra` includes both dependent and zero intra-sentential arguments.

$h^2 = (h_0^2, h_1^2)$ is a two-dimensional vector where h_0^2 is a probability that a word is the argument of the predicate and h_1^2 is a probability that a word is not the argument of the predicate.

$$\bar{x} = w_a \oplus w_f \oplus b_f \quad (15)$$

$$h^1 = \text{BiLSTM}(\bar{x}) \quad (16)$$

$$h^2 = \text{linear}(h^1) \quad (17)$$

$$h^3 = (f^m(h_0^2), g^m(h_1^2)) \quad (18)$$

$$p = \text{softmax}(h^3) \quad (19)$$

Equation (18) shifts the output probability by conditioning the distribution of argument types across the media sources.

(4) Voting This method determines its output based on the majority of the above three methods. When the decisions split, we select the decision with the highest probability.

(5) Mixture The last method combines the above three methods: Fine-tuning, Feature augmentation and Class probability shift into a single model.

5 Experiment

5.1 Setting

We evaluated our models on a Japanese balanced-corpus (BCCWJ). We divided the corpus into three portions: 70% for training, 10% for development, and 20% for testing. We trained each model for 10 epochs and used the best model in terms of accuracy in the test data for evaluation.

Hyper parameter The number of dimensions of word embeddings and POS embedding are 200 and 30, respectively. The dropout rate of BiLSTM is 0.2. The batch size is 32. Our models were optimized with Adam (Kingma and Ba, 2014) in which

\model	Baseline										Adaptation				
	OC	OY	Each-D		PM	PN	Small	All	Out-D	One-H	F-t	F-a	C-p	Vote	Mix
target\size	11,777	10,929	23,471	22,773	21,287	33,327	92,674	123,564	-	-	-	-	-	-	-
OC	61.2	53.6	51.6	58.8	58.6	58.3	66.9	68.5	62.4	66.9	67.7	65.6	61.4	69.6	67.2
OY	51.7	54.1	47.5	52.3	52.3	52.5	61.1	60.4	58.4	63.2	63.2	61.5	65.6	64.0	62.8
OW	54.4	57.2	74.3	66.1	64.9	67.4	74.5	76.5	70.6	76.8	73.0	76.1	70.2	77.6	77.2
PB	61.6	57.2	66.1	74.0	71.4	71.1	76.8	78.8	75.1	79.2	73.8	77.6	78.3	79.3	78.0
PM	55.0	52.3	66.2	75.5	72.9	72.8	76.3	76.4	75.1	78.8	74.4	78.7	78.0	80.0	77.4
PN	54.1	54.0	63.2	64.3	65.1	69.8	72.1	74.0	71.1	73.6	70.7	72.5	73.9	74.4	73.0

OC: QA tests, OY: blog texts, OW: white papers, PB: books, PM: magazines, PN: newspapers

Table 4: NOM accuracy

arg. type\ target # of predicates\ model	OC				OY				OW			
	Each-D	All	Vote	Mix	Each-D	All	Vote	Mix	Each-D	All	Vote	Mix
intra(dep)	73.5	85.5	86.0	84.1	62.0	78.0	79.5	76.2	79.2	83.7	82.6	81.7
intra(zero)	32.3	46.3	46.5	45.8	33.7	41.7	46.1	43.1	33.9	40.7	33.1	34.4
exo1	50.4	36.3	45.5	47.8	58.4	24.9	37.5	60.1	0.0	0.0	0.0	0.0
exo2	40.5	40.1	46.4	36.1	9.1	27.3	13.6	20.5	0.0	0.0	0.0	0.0
unknown	66.3	75.4	73.5	69.8	53.7	76.1	76.3	60.7	86.3	85.0	90.9	90.4

arg. type\ target # of predicates\ model	PB				PM				PN			
	Each-D	All	Vote	Mix	Each-D	All	Vote	Mix	Each-D	All	Vote	Mix
intra(dep)	79.7	86.2	86.8	85.6	77.6	86.2	85.5	82.7	76.3	84.0	83.3	81.5
intra(zero)	45.0	55.9	53.8	52.9	47.1	53.6	52.1	44.7	45.7	51.0	47.7	48.0
exo1	0.0	0.0	0.0	0.0	2.4	3.2	0.0	1.6	0.0	2.4	0.0	1.2
exo2	0.0	56.0	0.0	0.0	0.0	9.7	6.5	0.0	5.4	24.3	5.4	13.5
unknown	83.2	81.6	84.0	82.2	81.4	79.7	88.9	87.8	79.5	78.9	82.9	80.7

OC: QA tests, OY: blog texts, OW: white papers, PB: books, PM: magazines, PN: newspapers

Table 5: NOM accuracy details

α is 0.001, β is 0.9 and weight decay is 0. For fine-tuning, weight decay is 0.0001.

5.2 Result

Table 4 shows the result of each model for the nominative case. Due to the page limitation, we show only brief results of the nominative case. Appendix includes the detailed results of all cases.

For the Each-D models, the media source of the training data affected the accuracy. In general, the model worked better when the training examples and the test examples came from the same media source, i.e. the training data is in-domain.

The All model performed better than the Small model in all media sources. This reveals that the total amount of training examples is vital to improve the accuracy. These facts show that both the total amount of examples and the media source should be taken into consideration.

The All model performed better than the Each-D models in all media sources. This implies that on

top of the in-domain data, more training examples are effective even though they are out-domain.

However, removing the in-domain data from the training data degrades the performance even using large data. Although the training data of all Out-D models except for PN³ is larger than that of the Small model, the Small model worked better than the Out-D models. When the training data lacks the data from the target media source, the data size does not always compensate the discrepancy in domains between the training and test data.

The All model and the One-H model showed no significant difference in accuracy. It indicates that a one-hot vector about media does not work well.

The right-hand side in Table 4 shows the result of the domain adaptation methods: the Fine-tuning method (F-t), the Feature augmentation method (F-a), the Class probability shift method (C-p),

³The total amount of training examples in the Out-D model for PN is $123,564 - 33,327 = 90,237$.

media	Vote	All	Sentence
OC	exo2	unknown	<i>shiryokukaifuku-no yoi houhou-o oshiete kudasai.</i> of eyesight recovery good way _{ACC} tell _v please Please tell me a good way to recover my eyesight.
OY	exo1	unknown	<i>ippai mite eigo ganbarimasu.</i> a lot watch _v English will try hard I will watch a lot and try English hard.
PN	<i>Toyota</i> (Toyota)	unknown	<i>Toyotashintaisei happyou.</i> Toyota _{NOM} new structure _{ACC} announcement _v Toyota announces its new structure.

Table 6: Examples analyzed correctly by the domain adaptation

the Voting method (*Vote*) and the Mixture method (*Mix*). The *Vote* model worked better than the *All* model in all media sources. Comparing the *Vote* models with other domain adaptation models, the *C-p* model worked better than the *Vote* model in the *OY* texts. The number of training examples in the *OY* texts is the smallest among the six media sources. The data size could be the main reason for the low performance in *OY*.

Table 5 shows the accuracy of each argument type in each media source. According to Table 3, *exo2* and *exo1* in *OC* and *exo2* in *OY* are frequent types. Therefore analyzing these exophoric types correctly contributes to the total performance.

Table 5 shows the accuracy of these types improved. The domain adaptation is successful for resolving these types of exophora.

Table 6 shows examples analyzed correctly by the *Vote* model, but not by the *All* model. Target predicates are shown in bold type. The *OC* texts contain QA texts like dialogue. Therefore a hearer (*exo1*) tends to fill the *NOM* case of the predicate as in “*oshiete* (tell)” in the first example. The *OY* texts contain blog texts where speakers often write their experiences. In such cases, a speaker tends to fill the *NOM* case (the second example in Table 6). Unlike these two media sources, the *PN* texts contain newspaper articles. The case markers tend to be omitted particularly in their titles in which the first phrase in the title tends to fill the *NOM* case. Our model successfully learned this tendency as shown in the last example.

6 Conclusion

This paper proposed effective domain adaptation methods for Japanese PAS analysis in various do-

mains (media sources). We proposed an RNN-based model as well as five domain adaptation methods. The evaluation experiments with a Japanese balanced-corpus (BCCWJ) confirmed that the domain adaptation is effective for improving the performance of the analysis.

Acknowledgments

We appreciate Dr. Yuichiroh Matsubayashi and Dr. Ryohei Sasano for the discussion on our problem setting.

References

- Chen Chen and Vincent Ng. 2016. Chinese Zero Pronoun Resolution with Deep Neural Networks. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 778–788.
- Alex Graves, Santiago Fernández, and Jürgen Schmidhuber. 2005. Bidirectional LSTM Networks for Improved Phoneme Classification and Recognition. In *Proceedings of the 15th International Conference on Artificial Neural Networks: Formal Models and Their Applications - Volume Part II, ICANN’05*, pages 799–804, Berlin, Heidelberg. Springer-Verlag.
- Jan Hajič, Massimiliano Ciaramita, Richard Johansson, Daisuke Kawahara, Maria Antònia Martí, Lluís Màrquez, Adam Meyers, Joakim Nivre, Sebastian Padó, Jan Štěpánek, Pavel Straňák, Mihai Surdeanu, Nianwen Xue, and Yi Zhang. 2009. The CoNLL-2009 Shared Task: Syntactic and Semantic Dependencies in Multiple Languages. In *Proceedings of the Thirteenth Conference on Computational Natural Language*

- Learning: Shared Task*, CoNLL '09, pages 1–18, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Masatsugu Hangyo, Daisuke Kawahara, and Sadao Kurohashi. 2013. Japanese Zero Reference Resolution Considering Exophora and Author/Reader Mentions. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 924–934. Association for Computational Linguistics.
- Yuta Hayashibe, Mamoru Komachi, and Yuji Matsumoto. 2011. Japanese Predicate Argument Structure Analysis Exploiting Argument Position and Type. In *Proceedings of 5th International Joint Conference on Natural Language Processing*, pages 201–209. Asian Federation of Natural Language Processing.
- Luheng He, Kenton Lee, Mike Lewis, and Luke Zettlemoyer. 2017. Deep Semantic Role Labeling: What Works and What’s Next. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 473–483. Association for Computational Linguistics.
- Ryu Iida and Massimo Poesio. 2011. A Cross-Lingual ILP Solution to Zero Anaphora Resolution. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 804–813. Association for Computational Linguistics.
- Kenji Imamura, Ryuichiro Higashinaka, and Tomoko Izumi. 2014. Predicate-Argument Structure Analysis with Zero-Anaphora Resolution for Dialogue Systems. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 806–815, Dublin, Ireland. Dublin City University and Association for Computational Linguistics.
- 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, ACLShort '09*, pages 85–88, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Yuhko Kayama. 2003. L1 acquisition of Japanese zero pronouns: The effect of discourse factors. In *Proceedings of the 2003 Annual Conference of the Canadian Linguistic Association*, pages 109–120.
- Young-Bum Kim, Karl Stratos, and Ruhi Sarikaya. 2016. Frustratingly Easy Neural Domain Adaptation. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 387–396. The COLING 2016 Organizing Committee.
- Diederik P. Kingma and Jimmy Ba. 2014. Adam: A Method for Stochastic Optimization. *CoRR*, abs/1412.6980.
- Kikuo Maekawa, Makoto Yamazaki, Toshinobu Ogiso, Takehiko Maruyama, Hideki Ogura, Wakako Kashino, Hanae Koiso, Masaya Yamaguchi, Makiro Tanaka, and Yasuharu Den. 2014. Balanced Corpus of Contemporary Written Japanese. *Lang. Resour. Eval.*, 48(2):345–371.
- Yuichiroh Matsubayashi and Kentaro Inui. 2017. Revisiting the Design Issues of Local Models for Japanese Predicate-Argument Structure Analysis. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 128–133. Asian Federation of Natural Language Processing.
- Hiroki Ouchi, Hiroyuki Shindo, Kevin Duh, and Yuji Matsumoto. 2015. Joint Case Argument Identification for Japanese Predicate Argument Structure Analysis. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 961–970. Association for Computational Linguistics.
- Hiroki Ouchi, Hiroyuki Shindo, and Yuji Matsumoto. 2017. Neural Modeling of Multi-Predicate Interactions for Japanese Predicate Argument Structure Analysis. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1591–1600. Association for Computational Linguistics.
- Luz Rello, Ricardo Baeza-Yates, and Ruslan Mitkov. 2012. Elliphant: Improved Automatic

- Detection of Zero Subjects and Impersonal Constructions in Spanish. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics, EACL '12*, pages 706–715, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Ryohei Sasano and Sadao Kurohashi. 2011. A Discriminative Approach to Japanese Zero Anaphora Resolution with Large-scale Lexicalized Case Frames. In *IJCNLP*.
- Mike Schuster and Kuldip K. Paliwal. 1997. Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, 45:2673–2681.
- Tomohide Shibata, Daisuke Kawahara, and Sadao Kurohashi. 2016. Neural Network-Based Model for Japanese Predicate Argument Structure Analysis. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1235–1244. Association for Computational Linguistics.
- Mihai Surdeanu, Richard Johansson, Adam Meyers, Lluís Màrquez, and Joakim Nivre. 2008. The CoNLL-2008 Shared Task on Joint Parsing of Syntactic and Semantic Dependencies. In *Proceedings of the Twelfth Conference on Computational Natural Language Learning, CoNLL '08*, pages 159–177, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Masatoshi Suzuki, Koji Matsuda, Satoshi Sekine, Naoaki Okazaki, and Kentaro Inui. 2016. Neural Joint Learning for Classifying Wikipedia Articles into Fine-grained Named Entity Types. In *Proceedings of the 30th Pacific Asia Conference on Language, Information and Computation: Posters*, pages 535–544.
- Hirotoishi 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, EMNLP '08*, pages 523–532, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Haitong Yang, Tao Zhuang, and Chengqing Zong. 2015. Domain Adaptation for Syntactic and Semantic Dependency Parsing Using Deep Belief Networks. *Transactions of the Association for Computational Linguistics*, 3:271–282.
- Qingyu Yin, Yu Zhang, Weinan Zhang, and Ting Liu. 2017. Chinese Zero Pronoun Resolution with Deep Memory Network. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1309–1318. Association for Computational Linguistics.
- Katsumasa Yoshikawa, Masayuki Asahara, and Yuji Matsumoto. 2013. Jointly Extracting Japanese Predicate-Argument Relation with Markov Logic. *Journal of Natural Language Processing*, 20(2):251–271.
- Jie Zhou and Wei Xu. 2015. End-to-end learning of semantic role labeling using recurrent neural networks. In *ACL (1)*, pages 1127–1137. The Association for Computer Linguistics.

Appendix: Experimental results

\model	Baseline										Adaptation				
	OC	OY	Each-D		PM	PN	Small	All	Out-D	One-H	F-t	F-a	C-p	Vote	Mix
target\size	11,777	10,929	23,471	22,773	21,287	33,327	92,674	123,564	-	-	-	-	-	-	-
OC	61.2	53.6	51.6	58.8	58.6	58.3	66.9	68.5	62.4	66.9	67.7	65.6	61.4	69.6	67.2
OY	51.7	54.1	47.5	52.3	52.3	52.5	61.1	60.4	58.4	63.2	63.2	61.5	65.6	64.0	62.8
OW	54.4	57.2	74.3	66.1	64.9	67.4	74.5	76.5	70.6	76.8	73.0	76.1	70.2	77.6	77.2
PB	61.6	57.2	66.1	74.0	71.4	71.1	76.8	78.8	75.1	79.2	73.8	77.6	78.3	79.3	78.0
PM	55.0	52.3	66.2	75.5	72.9	72.8	76.3	76.4	75.1	78.8	74.4	78.7	78.0	80.0	77.4
PN	54.1	54.0	63.2	64.3	65.1	69.8	72.1	74.0	71.1	73.6	70.7	72.5	73.9	74.4	73.0

OC: QA tests, OY: blog texts, OW: white papers, PB: books, PM: magazines, PN: newspapers

Table 7: Result for nominative case (NOM) (accuracy)

\model	Baseline										Adaptation				
	OC	OY	Each-D		PM	PN	Small	All	Out-D	One-H	F-t	F-a	C-p	Vote	Mix
target\size	11,777	10,929	23,471	22,773	21,287	33,327	92,674	123,564	-	-	-	-	-	-	-
OC	83.4	79.0	78.2	81.7	80.1	80.1	83.9	85.3	83.6	85.9	85.3	84.8	84.4	86.1	85.9
OY	81.0	82.4	77.6	80.3	84.0	82.9	84.9	85.8	83.9	86.2	83.8	84.7	85.8	85.8	84.7
OW	65.4	64.9	79.8	73.1	72.4	74.3	81.1	82.1	78.4	81.7	77.9	81.8	76.2	82.9	80.6
PB	84.8	83.7	83.9	86.5	85.8	86.1	88.1	88.6	88.0	89.0	86.6	87.8	88.6	88.3	88.8
PM	80.9	81.2	80.3	82.9	84.2	83.8	85.8	86.4	86.2	85.8	83.7	86.0	85.7	86.3	85.4
PN	77.5	78.1	80.1	79.9	81.6	83.8	85.1	85.8	84.4	85.8	83.8	85.3	85.5	85.9	85.2

OC: QA tests, OY: blog texts, OW: white papers, PB: books, PM: magazines, PN: newspapers

Table 8: Result for accusative case (ACC) (accuracy)

\model	Baseline										Adaptation				
	OC	OY	Each-D		PM	PN	Small	All	Out-D	One-H	F-t	F-a	C-p	Vote	Mix
target\size	11,777	10,929	23,471	22,773	21,287	33,327	92,674	123,564	-	-	-	-	-	-	-
OC	86.8	84.6	84.0	85.9	85.9	85.6	88.7	88.7	87.3	88.9	87.5	87.9	87.8	89.5	88.7
OY	90.7	91.8	91.5	91.0	92.3	91.8	92.4	92.9	91.7	92.3	92.0	92.2	92.4	92.5	92.2
OW	87.6	87.5	90.5	88.8	88.7	89.2	90.7	91.2	90.4	90.9	89.6	90.8	88.6	91.0	90.1
PB	88.3	88.1	87.6	90.3	89.8	89.2	91.0	91.2	90.6	90.8	90.1	90.4	91.0	91.1	91.2
PM	88.3	88.2	87.6	89.0	90.6	90.0	91.4	91.3	91.3	90.8	90.0	90.4	91.1	91.2	91.3
PN	89.5	90.3	90.6	89.7	91.4	91.7	92.1	92.5	91.8	92.4	91.7	91.9	92.3	92.3	92.4

OC: QA tests, OY: blog texts, OW: white papers, PB: books, PM: magazines, PN: newspapers

Table 9: Result for dative case (DAT) (accuracy)

arg. type \ target # of predicates \ model	OC									
	Each-D	Small	All	Out-D	One-H	F-t	F-a	C-p	Vote	Mix
none	-	-	-	-	-	-	-	-	-	-
intra(dep)	73.5	82.5	85.5	82.9	85.4	83.6	84.0	80.8	86.0	84.1
intra(zero)	32.3	45.8	46.3	39.8	45.8	43.8	44.3	32.8	46.5	45.8
exo1	50.4	40.2	36.3	24.0	41.4	44.3	36.6	0.0	45.5	47.8
exo2	40.5	25.8	40.1	21.8	42.5	42.5	50.8	14.3	46.4	36.1
unknown	66.3	75.9	75.4	70.5	68.5	72.6	66.3	83.2	73.5	69.8
arg. type \ target # of predicates \ model	OY									
	Each-D	Small	All	Out-D	One-H	F-t	F-a	C-p	Vote	Mix
none	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
intra(dep)	62.0	77.6	78.0	77.1	78.3	79.7	76.7	79.1	79.5	76.2
intra(zero)	33.7	41.4	41.7	42.9	45.4	43.9	47.4	44.4	46.1	43.1
exo1	58.4	27.0	24.9	10.1	57.8	49.4	56.8	63.6	37.5	60.1
exo2	9.1	11.4	27.3	15.9	22.7	6.8	27.3	22.7	13.6	20.5
unknown	53.7	78.3	76.1	81.0	60.2	66.2	56.1	63.8	76.3	60.7
arg. type \ target # of predicates \ model	OW									
	Each-D	Small	All	Out-D	One-H	F-t	F-a	C-p	Vote	Mix
none	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
intra(dep)	79.2	77.8	83.7	77.3	81.5	73.2	80.3	80.2	82.6	81.7
intra(zero)	33.9	32.2	40.7	22.9	34.1	33.2	32.1	26.6	33.1	34.4
exo1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
exo2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0
unknown	86.3	88.1	85.0	83.9	89.8	88.0	89.9	79.5	90.9	90.4
arg. type \ target # of predicates \ model	PB									
	Each-D	Small	All	Out-D	One-H	F-t	F-a	C-p	Vote	Mix
none	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
intra(dep)	79.7	84.7	86.2	85.5	87.1	77.5	86.0	86.1	86.8	85.6
intra(zero)	45.0	51.1	55.9	51.3	54.4	45.2	52.8	54.5	53.8	52.9
exo1	0.0	14.3	0.0	14.3	0.0	0.0	0.0	0.0	0.0	0.0
exo2	0.0	16.0	56.0	24.0	0.0	0.0	0.0	20.0	0.0	0.0
unknown	83.2	80.8	81.6	74.1	83.4	85.6	80.4	81.3	84.0	82.2
arg. type \ target # of predicates \ model	PM									
	Each-D	Small	All	Out-D	One-H	F-t	F-a	C-p	Vote	Mix
none	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
intra(dep)	77.6	84.2	86.2	85.1	86.6	75.1	85.1	86.4	85.5	82.7
intra(zero)	47.1	51.3	53.6	51.0	52.9	44.2	53.6	51.6	52.1	44.7
exo1	2.4	2.4	3.2	5.5	0.8	1.6	0.8	0.0	0.0	1.6
exo2	0.0	9.7	9.7	9.7	3.2	12.9	12.9	12.9	6.5	0.0
unknown	81.4	82.0	79.7	78.5	85.0	87.7	85.8	83.9	88.9	87.8
arg. type \ target # of predicates \ model	PN									
	Each-D	Small	All	Out-D	One-H	F-t	F-a	C-p	Vote	Mix
none	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
intra(dep)	76.3	81.0	84.0	82.5	83.7	78.3	80.6	83.5	83.3	81.5
intra(zero)	45.7	46.1	51.0	47.9	47.9	43.1	45.7	49.7	47.7	48.0
exo1	0.0	6.1	2.4	7.3	1.2	0.0	8.5	0.0	0.0	1.2
exo2	5.4	2.7	24.3	2.7	5.4	0.0	21.6	13.5	5.4	13.5
unknown	79.5	79.8	78.9	74.2	80.0	81.2	81.4	79.8	82.9	80.7

Table 10: Detailed result for nominative case (NOM) (accuracy)

arg. type \ target # of predicates \ model	OC									
	Each-D	Small	All	Out-D	One-H	F-t	F-a	C-p	Vote	Mix
none	95.8	96.8	96.4	96.7	95.8	94.4	95.4	96.8	96.0	95.2
intra(dep)	73.1	76.7	80.7	79.7	82.6	79.3	79.3	82.8	81.1	82.1
intra(zero)	27.7	27.0	34.0	28.3	23.9	30.8	28.3	39.0	27.7	30.2
exo1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	20.0
exo2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	25.0	0.0
unknown	58.4	50.3	55.7	41.9	64.1	70.9	61.7	38.6	66.6	65.5
arg. type \ target # of predicates \ model	OY									
	Each-D	Small	All	Out-D	One-H	F-t	F-a	C-p	Vote	Mix
none	93.5	93.5	92.4	91.7	94.0	93.2	94.4	93.0	93.6	93.4
intra(dep)	70.3	76.9	82.4	76.5	81.3	73.6	74.5	82.3	79.7	77.8
intra(zero)	22.6	23.3	32.3	33.1	24.1	21.1	27.1	32.3	29.3	27.1
exo1	0.0	0.0	0.0	0.0	16.7	0.0	33.3	0.0	0.0	0.0
exo2	-	-	-	-	-	-	-	-	-	-
unknown	39.0	57.9	59.8	57.2	57.9	57.2	48.4	50.9	57.2	48.4
arg. type \ target # of predicates \ model	OW									
	Each-D	Small	All	Out-D	One-H	F-t	F-a	C-p	Vote	Mix
none	90.8	91.2	90.4	92.5	91.6	93.3	90.9	90.3	92.4	90.2
intra(dep)	86.9	88.9	89.1	83.1	88.5	81.2	88.0	83.6	89.3	88.3
intra(zero)	27.6	26.5	28.9	18.9	24.7	25.0	27.8	17.4	29.3	32.3
exo1	-	-	-	-	-	-	-	-	-	-
exo2	-	-	-	-	-	-	-	-	-	-
unknown	34.8	39.2	51.8	36.8	46.5	31.6	51.8	20.9	49.7	37.3
arg. type \ target # of predicates \ model	PB									
	Each-D	Small	All	Out-D	One-H	F-t	F-a	C-p	Vote	Mix
none	96.5	96.9	96.9	97.4	96.3	96.8	95.4	96.7	96.9	97.2
intra(dep)	80.6	86.9	87.5	84.5	88.1	80.0	87.3	88.2	86.6	86.4
intra(zero)	26.0	23.0	25.3	21.3	28.3	23.3	32.0	28.3	29.7	26.7
exo1	-	-	-	-	-	-	-	-	-	-
exo2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
unknown	41.8	39.3	44.7	44.0	54.4	46.5	42.8	39.0	39.3	49.1
arg. type \ target # of predicates \ model	PM									
	Each-D	Small	All	Out-D	One-H	F-t	F-a	C-p	Vote	Mix
none	95.2	94.4	94.5	93.2	94.2	95.8	95.7	93.2	96.0	94.4
intra(dep)	84.9	88.8	90.0	90.3	88.7	81.8	87.7	90.3	88.4	89.0
intra(zero)	27.2	24.1	31.3	33.8	26.0	23.3	35.7	27.4	30.2	27.7
exo1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
exo2	-	-	-	-	-	-	-	-	-	-
unknown	29.6	47.8	43.9	51.1	48.6	34.9	32.4	48.6	36.9	37.2
arg. type \ target # of predicates \ model	PN									
	Each-D	Small	All	Out-D	One-H	F-t	F-a	C-p	Vote	Mix
none	90.9	91.7	91.7	91.3	93.3	93.1	93.0	91.8	93.0	91.4
intra(dep)	84.4	86.2	87.2	84.7	86.5	82.7	85.6	87.4	87.0	86.4
intra(zero)	23.1	22.3	25.2	21.2	19.1	19.1	23.6	25.2	24.1	21.2
exo1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
exo2	-	-	-	-	-	-	-	-	-	-
unknown	37.7	45.8	50.2	47.4	44.7	24.9	35.1	41.4	37.7	51.6

OC: QA tests, OY: blog texts, OW: white papers, PB: books, PM: magazines, PN: newspapers

Table 11: Detailed result for accusative case (ACC) (accuracy)

arg. type \ target # of predicates \ model	OC									
	Each-D	Small	All	Out-D	One-H	F-t	F-a	C-p	Vote	Mix
none	97.3	98.2	98.8	99.3	98.8	98.2	96.6	99.2	98.2	97.0
intra(dep)	65.9	75.8	71.7	67.9	73.9	72.8	75.0	74.7	78.0	77.5
intra(zero)	16.7	16.7	15.3	11.1	12.5	16.7	15.3	15.3	16.7	19.4
exo1	54.9	51.0	54.9	0.0	58.8	60.8	54.9	0.0	60.8	64.7
exo2	68.2	9.1	36.4	22.7	50.0	45.5	0.0	31.8	63.6	63.6
unknown	15.8	26.3	22.1	14.7	20.5	4.2	37.9	11.1	26.3	29.5
arg. type \ target # of predicates \ model	OY									
	Each-D	Small	All	Out-D	One-H	F-t	F-a	C-p	Vote	Mix
none	98.1	97.0	98.0	96.1	97.8	98.2	97.2	97.5	97.6	97.5
intra(dep)	60.9	78.2	73.6	74.7	70.9	62.8	74.7	73.6	75.1	69.4
intra(zero)	7.4	22.2	11.1	18.5	11.1	14.8	18.5	18.5	18.5	22.2
exo1	0.0	7.1	0.0	7.1	7.1	0.0	0.0	7.1	0.0	14.3
exo2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
unknown	6.4	7.9	11.1	27.0	0.0	0.0	1.6	3.2	0.0	7.9
arg. type \ target # of predicates \ model	OW									
	Each-D	Small	All	Out-D	One-H	F-t	F-a	C-p	Vote	Mix
none	98.7	98.1	98.8	98.5	98.4	99.3	98.0	95.3	98.7	96.3
intra(dep)	59.4	64.1	64.8	59.9	64.1	47.4	68.0	67.7	63.6	68.8
intra(zero)	7.8	10.3	8.6	8.6	11.2	5.2	9.5	11.2	8.6	12.9
exo1	-	-	-	-	-	-	-	-	-	-
exo2	-	-	-	-	-	-	-	-	-	-
unknown	2.9	11.2	5.9	2.4	4.7	0.0	2.4	4.1	2.9	21.8
arg. type \ target # of predicates \ model	PB									
	Each-D	Small	All	Out-D	One-H	F-t	F-a	C-p	Vote	Mix
none	97.2	97.4	98.4	98.1	97.7	96.4	96.7	97.1	97.6	97.3
intra(dep)	76.3	81.4	76.5	75.8	78.3	79.0	82.6	82.1	81.3	80.2
intra(zero)	7.8	11.4	10.6	7.8	9.9	13.5	10.6	13.5	13.5	12.8
exo1	-	-	-	-	-	-	-	-	-	-
exo2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
unknown	14.2	8.5	6.3	3.4	6.3	15.9	2.8	11.4	5.1	20.5
arg. type \ target # of predicates \ model	PM									
	Each-D	Small	All	Out-D	One-H	F-t	F-a	C-p	Vote	Mix
none	97.2	96.6	97.5	97.0	97.2	98.3	96.5	96.0	97.6	96.8
intra(dep)	74.7	83.4	77.8	79.4	77.0	65.7	78.1	81.5	78.0	79.8
intra(zero)	15.7	16.5	9.6	20.0	10.4	4.4	12.2	13.9	7.8	17.4
exo1	0.0	0.0	25.0	25.0	25.0	0.0	0.0	25.0	0.0	0.0
exo2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
unknown	3.2	8.9	11.3	6.5	0.8	0.0	2.4	31.5	0.8	19.4
arg. type \ target # of predicates \ model	PN									
	Each-D	Small	All	Out-D	One-H	F-t	F-a	C-p	Vote	Mix
none	97.9	96.8	97.9	96.9	97.9	98.4	97.2	96.9	97.7	97.1
intra(dep)	67.0	77.9	73.3	75.8	73.0	63.3	73.5	78.5	73.5	77.7
intra(zero)	7.5	9.0	7.5	6.0	6.7	3.7	9.0	9.7	6.7	9.0
exo1	-	-	-	-	-	-	-	-	-	-
exo2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
unknown	4.4	8.1	9.6	2.9	1.5	2.2	1.5	8.8	4.4	9.6

OC: QA tests, OY: blog texts, OW: white papers, PB: books, PM: magazines, PN: newspapers

Table 12: Detailed result for dative case (DAT) (accuracy)