Predicate Argument Structure Analysis using Partially Annotated Corpora

Koichiro Yoshino, Shinsuke Mori, Tatsuya Kawahara
School of Informatics, Kyoto University
Yoshida-Honmachi, Sakyo-ku, Kyoto 606-8501, Japan
yoshino@ar.media.kyoto-u.ac.jp

Abstract

We present a novel scheme of predicate argument structure analysis that can be trained from partially annotated corpora. In order to allow partial annotation, this semantic role labeler does not require word dependency information. The advantage of partial annotation is that it allows for smooth domain adaptation of training data and improves the adaptability to a variety of domains.

1 Introduction

The predicate-argument (P-A) structure is one of the most fundamental and important representations in linguistics (Fillmore, 1968). Many applications use P-A structure as a component, for example, QA systems (Shen and Lapata, 2007), text mining systems (Wang and Zhang, 2009), and a spoken dialogue systems (Yoshino et al., 2011).

P-A structure analysis is regarded as a task of semantic role labeling (SRL). A semantic role represents a meaning of the components in P-A structure (i.e. Propbank (Palmer et al., 2005), FrameNet (Baker et al., 1998), and NAIST Text Corpus (NTC) (Iida et al., 2007b)). Traditional P-A structure analyzers estimate the semantic role labels for an input sentence by referring to a model trained on data annotated with not only semantic role labels but also dependency labels (Surdeanu et al., 2008; Hajič et al., 2009). Most of the previous approaches to P-A structure analysis assume full annotation for P-A structures and the lower layer labels: word boundaries, parts of speech (POS), and dependencies. Given a corpus fully annotated with them, the structural prediction approach was shown to be effective (Watanabe et al., 2010). However, this pre-annotation incurs high annotation costs which prevent us from adapting the analyzer to new domains. Having training data that are representative of a domain is essential for constructing a robust semantic role labeler (Pradhan et al., 2008) because the important information structures are specific to each domain (R.Grishman, 2003). Fully annotated corpus in target domain is not available in realistic cases, and it is difficult to apply the current supervised approaches to a new domain.

When annotating only the domain-specific area, the use of a partially annotated corpus (Tsuboi et al., 2008; Sassano and Kurohashi, 2010) that allows incomplete annotations improves accuracy efficiently and reduce the number of annotations. The pointwise approach (Neubig and Mori, 2010) enables efficient use of such incomplete language resources in word segmentation tasks and requires only partial annotations for the relevant tasks and lower layer annotations on which they depend. We design a new P-A structure analysis method that enables us to directly estimate semantic role labels by referring to a model that is trained from a corpus that includes only partially annotated tag information without word dependencies.

2 Predicate argument structure analysis

In this section, we give a brief explanation of P-A structure and its problems. Then, we describe the typical method of structural prediction for this task based on supervised machine learning.

2.1 Predicate-argument (P-A) structure

A predicate-argument (P-A) structure is a relationship between a verbal expression and its arguments, such as the subject, the direct object, and the indirect object. Predicate P in a document D has arguments $A_1, A_2, ..., A_n$ that have a semantic role $S_1, S_2, ..., S_n$. The notion we used is defined in NAIST Text Corpus (NTC) (Iida et al., 2007b), Japanese text corpora annotated with coreference and P-A relations, which include annotations of subject, direct object, and indirect object. Every case has a property of depend or zero, and these P-A structure relations are annotated not only predicates, but also event nouns (Komachi et al., 2007).

Figure 5 shows an example of P-A structures in the NTC. These tags have two properties, one is depend or zero, the other is *intra* or *inter*. depend or zero indicates whether or not the argument has a dependency on the predicate, and *intra*

International Joint Conference on Natural Language Processing, pages 957–961,
Nagoya, Japan, 14-18 October 2013.

or *inter* indicates whether or not the argument and the predicate exist in the same sentence. NTC also includes annotations of coreference, which we converted into P-A structure tags. As shown in **Figure 1**, P-A structure is located in a higher layer of linguistics that approaches natural language understanding (NLU), and this structure depends on some more basic but much more frequent linguistic phenomena: word boundaries, part of speech (POS), and word dependencies.

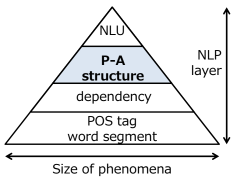

Figure 1: Size of linguistic phenomena.

2.2 Typical solution

The typical solution divides the P-A structure prediction into two problems: semantic role labeling (SRL) (Johansson and Nugues, 2008; Björkelund et al., 2009) and zero-anaphora resolution (Iida et al., 2007a; Sasano and Kurohashi, 2009).

The typical approach requires three preprocessing steps: word segmentation, POS tagging, and dependency parsing. After the preprocessing, the task of SRL improves assigning semantic role labels to the edges in word dependencies. A semantic role labeler performs two tasks: predicate sense estimation, and SRL. Zero-anaphora resolution is treated as an independent problem from the SRL task in the previous research. The zero-anaphora problem is caused by the ellipsis of shared words, and it is a gap in a sentence that has an anaphoric function (Iida et al., 2007a). Some semantic relationships exist in which there is no dependency relationship between their arguments and predicates; this is called zero anaphora.

The task of P-A structure analysis goes beyond the syntactic problem and comes down to a semantic problem to fill in the words that are semantically omitted. Various special approaches can be applied after SRL to solve this problem (Sasano and Kurohashi, 2009; Iida and Poesio, 2011; Hayashibe et al., 2011). Some approaches adopt a *Salience Reference List* (Nariyama, 2002) based on the 1-best argument decision model.

2.3 Open problems in P-A structure analysis

Existing approaches require full annotation of word boundaries, POS tags, and word dependencies to use them as features. Most previous approaches to P-A structure tasks assume full annotation for these lower layers. However, the number of linguistic phenomena decreases as we go higher up the NLP layers as shown in Figure 1. Thus, to prepare only one training example for an existing

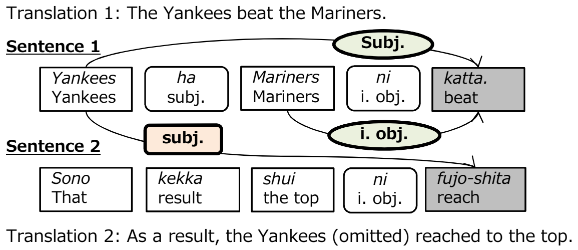

Figure 2: Example of training data made from partially annotated corpus.

P-A structure analyzer, we need to annotate the entire document. To make it worse, these kinds of annotations are costly and difficult for untrained annotators. This difficulty interferes with efficient language resource preparation and reduces domain portability. However, the accuracy of P-A structure analysis increases in accordance with the data size. This indicates that we can realize an improvement just by easily preparing more training data for the target domain document.

3 Partial annotation for P-A structures

Partial annotation allows annotators to focus on efficient examples in the target domain document, and to maximize the cost-effectiveness of annotation. For automatic word segmentation and POS tagging, the scheme allows partial annotation of corpus (Tsuboi et al., 2008; Neubig and Mori, 2010; Neubig et al., 2011) and achieves high accuracy and domain portability though annotation of domain-specific areas. Neubig et al. (2011) report that a comparable accuracy to a CRF-based sequential labeling method can be achieved without referring to the estimated labels for unlabeled words. They call this method a pointwise approach. Even with the pointwise assumption, we can estimate labels as accurately as sequential labeling just by referring to the appropriate features.

We design a P-A structure analyzer that directly estimates the semantic role labels by referring to a model that is trained from a corpus. It includes only partially annotated POS tags but not with dependency information for the following reasons. Automatic estimation of POS tags achieves high accuracy in domain adaptation cases, and the annotation cost is small (Neubig et al., 2011), but the accuracy for dependency parsing (Flannery et al., 2011; Sassano and Kurohashi, 2010) is not sufficiently high. However, handcraft annotation cost of dependency is so high, and it disturbs rapid preparation of annotation data.

We show an example of a partially annotated corpus in **Figure 2**. The annotation of "reach" is incomplete, and the information that can be referred to by an analyzer is the fully annotated word boundaries, POS tags, and partially annotated P-A tags. Word boundaries and POS tags are output by

Table 1: Features of SRL: w_p is a predicate, w_a is an argument candidate, t_i is the POS tag of w_i.

type	feature
word 1-gram	w_{p-3},w_{p-2},w_{p-1},w_p,w_{p+1},w_{p+2},w_{p+3}, w_{a-3},w_{a-2},w_{a-1},w_a,w_{a+1},w_{a+2},w_{a+3}
word 2-gram	$w_{p-1}w_p$, w_pw_{p+1},$w_{a-1}w_a$, w_aw_{a+1}
word 3-gram	$w_{p-1}w_pw_{p+1}$, $w_{a-1}w_aw_{a+1}$
POS 1-gram	t_{p-3},t_{p-2},t_{p-1},t_p,t_{p+1},t_{p+2},t_{p+3}, t_{a-3},t_{a-2},t_{a-1},t_a,t_{a+1},t_{a+2},t_{a+3}
POS 2-gram	$t_{p-1}t_p$, t_pt_{p+1},$t_{a-1}t_a$, t_at_{a+1}
POS 3-gram	$t_{p-1}t_pt_{p+1}$,$t_{a-1}t_at_{a+1}$
pairwise	Pairs of POS tags located -2 – +2. Pairs of arg candidate and pred.
distance	Number of pred between the candidate and preds
binary	(1) Closest candidate that has target particle on the right side or not. (2) First candidate that has target particle on the right side or not. (3) The predicate has a slot of target semantic case or not.

a domain-adapted morphological analyzer, and the annotator tags three P-A tags.

4 Pointwise P-A structure analysis

In our proposed scheme, syntactic ambiguity resolution and predicate sense disambiguation are not used, in order to achieve easy adaptation. We propose two sequential processes for P-A structure analysis that is trained from partially annotated corpora. Following the discussion in Section 3, we do not assume the dependency structures.

4.1 Case existence detection

The first step in the proposed sequential analysis is case existence estimation. The given semantic cases differ according to the type of the predicate. This predicate and semantic case behavior strongly affects the SRL task.

The oracle of the case existence is used for SRL features. For example, the predicate "bet" in Figure 5 contains information indicating that the predicate has two kinds of argument: "subject, zero" and "direct object depend." We assume that the case existence for each predicate can be estimated with case frames (Kawahara and Kurohashi, 2006). A case frame is a set of a predicate and its potential arguments. It is known that the case frames contribute to the P-A structure analysis performance (Sasano et al., 2008).

4.2 SRL and zero-anaphora resolution

The second step is SRL that includes zero-anaphora resolution. We handle the problem with a direct approach for SRL that is redefined as a binary classification problem for the pair of an argument candidate and a predicate. Labeled pairs of argument (arg) and predicate (pred) are used as positive training example and unlabeled pairs are used as negative training example. In the example of Figure 5, the pair of "fate" and "party" is a positive example Y, and pairs of "fate" and other candidates are negative examples N.

The features used for classification are listed in **Table 1**. We use simple n-gram features based on words and POS tags. The pairwise features are POS pairs located at positions from -2 to +2, and pairs of the predicate and the argument candidates. The distance between the argument candidate and the predicate is used as a feature. We used the number of predicates between the predicate and the argument candidate as this feature. Binary features (1) and (2) are based on a previous study on "Centering" theory (Grosz et al., 1995). In this theory, subjects are frequently omitted, and the first candidate tends to be a subject. By contrast, objects are not omitted, and the last candidate tends to be an object. To apply the theory to a pointwise approach, we define features that are independent of syntactic structure. Finally, the result of the processing described above is used as a binary feature (3).

4.3 Issues in partial annotation

Two problems arise in applying the classifier to a partially annotated corpus without dependency. First, in existing studies P-A structure analysis leverages a property of the P-A tags depend and zero (Iida et al., 2007b). Here, depend represents that the P-A tag is added on the edge of dependency, and zero means the pair of the predicate and argument does not have a relationship of dependency (=zero anaphora). However, it is impossible to use dependency information in our framework, and the attribution makes it difficult to detect the property of the P-A tag. To cope with this problem, we use sentence boundaries, which are trivial in unlabeled documents, for grouping the training set. The other problem is how to create training examples from incomplete annotations. To allow the incomplete annotation perfectly, we incorporate positive examples that are clearly annotated.

5 Evaluations

We conducted three experiments to evaluate the proposed method: SRL, corpus size discrimination, and domain adaptation.

5.1 Experimental settings

We use the NTC (Iida et al., 2007b) which is annotated with P-A relations and coreferences. The NTC is constructed from Japanese newspaper articles, and has two domains: news and editorials. In the NTC, there are three different types of annotation on pairs of predicates and their arguments: subject, direct object, and indirect object. Every tag has a property of depend or zero. The

Table 2: Results of P-A analysis (with case frame, using the property of depend and zero).

	role label	prec.	recall	F
dep.	subject	0.747	0.754	0.750
	d. obj.	0.908	0.930	0.919
	i. obj.	0.953	0.947	0.950
	total	0.839	0.849	0.844
	total (w.o. feat. (3))	0.744	0.683	0.712
zero	subject	0.305	0.120	0.172
	d. obj.	0.560	0.212	0.307
	i. obj.	0.402	0.127	0.192
	total	0.402	0.127	0.192
	total (w.o. feat. (3))	0.251	0.115	0.157
total		0.580	0.321	0.413
cf. (zero)	subject	0.265	0.302	0.282
	d. obj.	0.092	0.129	0.107
	i. obj.	0.048	0.041	0.044

Table 3: Results of P-A analysis (with case frame, using the property of *intra* and *inter*).

	role label	prec.	recall	F
intra	subject	0.624	0.520	0.567
	d. obj.	0.841	0.809	0.825
	i. obj.	0.868	0.807	0.836
	total	0.730	0.646	0.686
inter	subject	0.311	0.118	0.171
	d. obj.	0.320	0.048	0.083
	i. obj.	0.329	0.085	0.135
	total	0.312	0.111	0.164
total		**0.602**	**0.366**	**0.455**
cf. (*inter*)	subject	0.221	0.273	0.244
	d. obj.	0.050	0.101	0.066
	i. obj.	0.030	0.023	0.026

NTC has lower layer annotations: word boundaries, POS, and segment-based dependencies. We used the word segments and POS tags as-is, and constructed P-A classifiers. We evaluated the proposed SRL in the newspaper article domain. We used linear support vector machine (SVM) (Fan et al., 2008) with the one-versus-rest method, by using the features described in Table 1.

5.2 Evaluation of SRL

The results using 5-fold cross validation are listed in **Table 2** and **3**. Evaluations that are classified with the existing depend and zero property are given in Table 2. Classifiers used in "w.o. feat. (3)" do not refer to case existence features (the binary feature (3) in Table 1). We can see that case frames play a large role in improving the labeling accuracy. This depend and zero property is based on the dependency, which cannot be referred to in the pointwise approach. As an alternative, we used sentence boundaries for the tag classification and Table 3 shows the result. The bottom "cf." rows in Tables 2 and 3 are the result of the previous work (Sasano and Kurohashi, 2011) for comparison[1]. In the comparison, the accuracies of our work are comparable to the accuracies of the previous work. By comparing the total F measure in Table 2 (0.413) and that in Table 3 (0.455), we can say that this *intra* and *inter* property works better than the depend and zero property in our pointwise classifier.

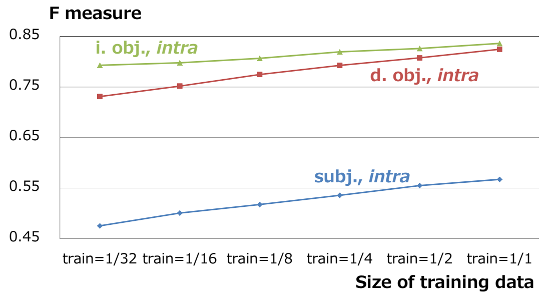

Figure 3: Effect of corpus size in *intra* case.

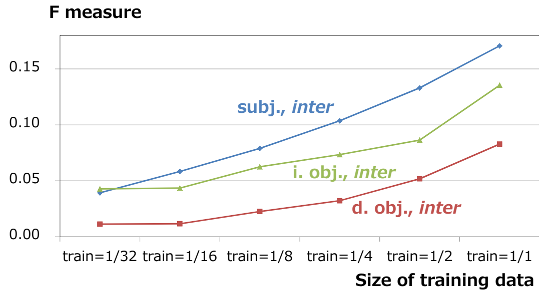

Figure 4: Effect of corpus size in *inter* case.

5.3 Effect of corpus size

We show the relationship between the training corpus size and the accuracy in **Figures 3** and **4**. The horizontal axes of these graphs are the log-scaled corpus size. The graphs show that P-A structure analysis accuracy increases linearly in proportion to the log-scaled data size and do not saturate. This result supports our framework of efficient resource usage.

6 Conclusion

We presented a novel scheme of P-A structure analysis based on the pointwise approach that makes it possible to use partially annotated corpora. This paper can be seen as an extension of the pointwise approach to a higher NLP layer that allows us to concentrate annotation work on the focused task. The results indicated that our scheme reduces the cost of constructing language resource and makes it easy to adapt the P-A structure analyzer while maintaining comparable accuracy to current analysis frameworks.

In future work, we plan to evaluate our pointwise P-A resolution method in the domain adaptation case in terms of personal costs of annotation and investigate improving accuracy by using other estimated information.

[1]Sasano and Kurohashi discussed this task, but the article is written in Japanese. They evaluated the accuracy for zero anaphora in two models: *intra* and *inter*, and we calculated the weighted mean of them for fair comparison.

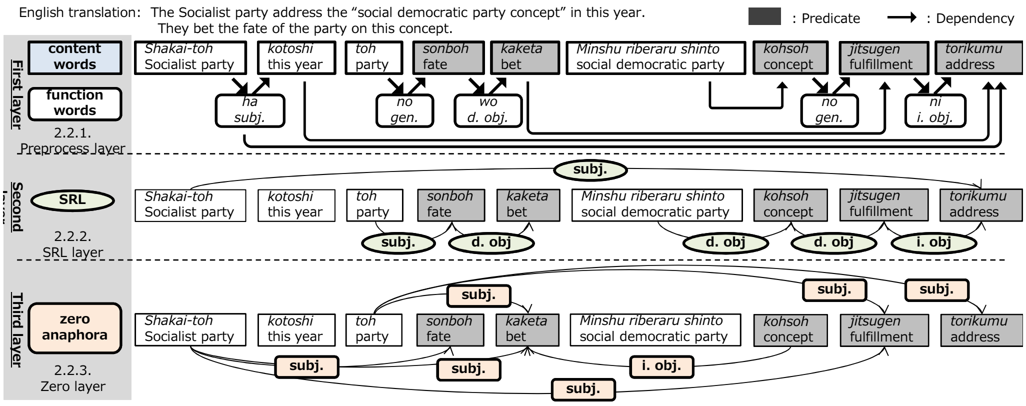

Figure 5: Example of P-A structure analysis.

A Figure 5 shows an example of P-A

References

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The berkeley framenet project. In *Proc. ACL-COLING*, pages 86–90.

Anders Björkelund, Love Hafdell, and Pierre Nugues. 2009. Multilingual semantic role labeling. In *Proc. CoNLL: Shared Task*, pages 43–48.

Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. 2008. Liblinear: A library for large linear classification. *Journal of Machine Learning Research*, 9(4):1871–1874.

Charles J. Fillmore. 1968. The case for case. In E. Bach and R. Harms, editors, *Universals in Linguistic Theory*.

Daniel Flannery, Yusuke Miyao, Graham Neubig, and Shinsuke Mori. 2011. Training dependency parsers from partially annotated corpora. In *Proc. IJCNLP*, pages 776–784.

Hagen Fürstenau and Mirella Lapata. 2009. Semi-supervised semantic role labeling. In *Proc. EACL*, pages 220–228.

Barbara J. Grosz, Scott Weinstein, and Aravind K. Joshi. 1995. Centering: a framework for modeling the local coherence of discourse. *Computational Linguistics*, 21(2):203–225.

Jan Hajič et al. 2009. The conll-2009 shared task: syntactic and semantic dependencies in multiple languages. In *Proc. CoNLL: Shared Task*, CoNLL '09, pages 1–18.

Yuta Hayashibe, Mamoru Komachi, and Yuji Matsumoto. 2011. Japanese predicate argument structure analysis exploiting argument position and type. In *Proc. IJCNLP*, pages 201–209.

Ryu Iida and Massimo Poesio. 2011. A cross-lingual ilp solution to zero anaphora resolution. In *Proc. ACL-HLT*, pages 804–813.

Ryu Iida, Kentaro Inui, and Yuji Matsumoto. 2007a. Zero-anaphora resolution by learning rich syntactic pattern features. *ACM Transactions on Asian Language Information Processing (TALIP)*, 6(4):12:1–12:22.

Ryu Iida, Mamoru Komachi, Kentaro Inui, and Yuji Matsumoto. 2007b. Annotating a Japanese text corpus with predicate-argument and coreference relations. In *Proc. the Linguistic Annotation Workshop*, pages 132–139.

Richard Johansson and Pierre Nugues. 2008. Dependency-based semantic role labeling of PropBank. In *Proc. EMNLP*, pages 69–78.

Daisuke Kawahara and Sadao Kurohashi. 2006. A fully-lexicalized probabilistic model for japanese syntactic and case structure analysis. In *Proc. HLT-NACCL*, pages 176–183.

Daisuke Kawahara, Sadao Kurohashi, and Koiti Hasida. 2002. Construction of a Japanese relevance-tagged corpus. In *Proc. LREC*, pages 2008–2013.

Mamoru Komachi, Ryu Iida, Kentaro Inui, and Yuji Matsumoto. 2007. Learning based argument structure analysis of event-nouns in japanese. In *Proc. PACLING*, pages 120–128.

Shinsuke Mori and Graham Neubig. 2011. A pointwise approach to pronunciation estimation for a tts front-end. In *Proc. INTERSPEECH*, pages 2181–2184.

Shigeko Nariyama. 2002. Grammar for ellipsis resolution in japanese. In *Proc. TMI*, pages 135–145.

Graham Neubig and Shinsuke Mori. 2010. Word-based partial annotation for efficient corpus construction. In *Proc. LREC*.

Graham Neubig, Yosuke Nakata, and Shinsuke Mori. 2011. Pointwise prediction for robust, adaptable Japanese morphological analysis. In *Proc. ACL*, pages 529–533.

Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. The proposition bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31(1):71–106.

Sameer S. Pradhan, Wayne Ward, and James H. Martin. 2008. Towards robust semantic role labeling. *Computational Linguistics*, 34(2):289–310, jun.

R.Grishman. 2003. Discovery methods for information extraction. In *Proc. ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition*, pages 243–247.

Ryohei Sasano and Sadao Kurohashi. 2009. A probabilistic model for associative anaphora resolution. In *Proc. EMNLP*, pages 1455–1464.

Ryohei Sasano and Sadao Kurohashi. 2011. A discriminative approach to japanese zero anaphora resolution with large-scale case frame. *Journal of Information Processing (in Japanese)*, 52(12):3328–3337.

Ryohei Sasano, Daisuke Kawahara, and Sadao Kurohashi. 2008. A fully-lexicalized probabilistic model for japanese zero anaphora resolution. In *Proc. COLING*, pages 769–776.

Manabu Sassano and Sadao Kurohashi. 2010. Using smaller constituents rather than sentences in active learning for japanese dependency parsing. In *Proc. ACL*, pages 356–365.

Dan Shen and Mirella Lapata. 2007. Using semantic roles to improve question answering. In *Proc. EMNLP-CoNLL*, pages 12–21.

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 *Proc. CoNLL: Shared Task*, pages 159–177.

Ivan Titov and Alexandre Klementiev. 2012. Semi-supervised semantic role labeling: Approaching from an unsupervised perspective. In *Proc. COLING*, pages 2635–2652.

Yuta Tsuboi, Hisashi Kashima, Hiroki Oda, Shinsuke Mori, and Yuji Matsumoto. 2008. Training conditional random fields using incomplete annotations. In *Proc. COLING*, pages 897–904.

Rui Wang and Yi Zhang. 2009. Recognizing textual relatedness with predicate-argument structure. In *Proc. EMNLP*, pages 784–792.

Yotaro Watanabe, Masayuki Asahara, and Yuji Matsumoto. 2010. A structured model for joint learning of argument roles and predicate senses. In *Proc. ACL*, pages 98–102.

Koichiro Yoshino, Shinsuke Mori, and Tatsuya Kawahara. 2011. Spoken dialogue system based on information extraction using similarity of predicate argument structures. In *Proc. SIGDIAL*, pages 59–66.

Koichiro Yoshino, Shinsuke Mori, and Tatsuya Kawahara. 2012. Language modeling for spoken dialogue system based on filtering using predicate-argument structures. In *Proc. COLING*, pages 2993–3002.