A case study of comparison of several methods for corpus-based speech intention identification

Kazutaka Shimada **Kaoru Iwashita** **Tsutomu Endo**
Department of Artificial Intelligence,
Kyushu Institute of Technology
680-4 Iizuka Fukuoka 820-8502 Japan
{shimada, k_iwa, endo}@pluto.ai.kyutech.ac.jp

Abstract

Dialogue analysis is one of the most important tasks for human-machine interaction. It is important for dialogue systems to recognize the intention of an utterance and the intentional structure of a discourse. If a system identifies the intentions of each utterance in a dialogue correctly, it can recognize the structure of the dialogue. This paper describes methods for intention identification of an utterance. In this paper we apply two methods to the task and evaluate the performance. The 1st method is based on a similarity measure between an input utterance and utterances in the corpus (Case examples). We compare several similarity measures in the experiment. The 2nd method is based on the Maximum Entropy (ME) method. We compare our methods with related work. In the experiment, the ME method produced the best performance. For the coverage, the similarity based method with the inner product outperformed the ME method. In addition, we verify the effect of dependency relations between words and context information for this intention identification task.

1 Introduction

Dialogue analysis is one of the most important tasks for human-machine interaction. Speech understanding systems have been developed to practical use recently. However, speech understanding is just one component of dialogue understanding systems. It is important for dialogue systems to recognize the intention of an utterance and the intentional structure of a discourse. Recognition of them leads to generation of more appropriate model of plans for discourse and a solution of a problem (Higashinaka et al., 2003, 2005). If a system estimates the intentions of each utterance in a dialogue correctly, it can recognize the structure of the dialogue. Some researchers have reported rule-based or example-based intention understanding methods (Kimura et al., 1998; Matsubara et al., 2002; Irie et al., 2003, 2004; Inui et al., 2003). Kimura et al. (1998) have reported a rule-based method for intention understanding. In general, constructing rules for the method, however, is costly. Irie et al. (2003) have proposed an example-based method for the task. They used a similarity measure and a sequence of utterances for the intention understanding. Inui et al. (2003) have proposed a method for classification of open-ended questionnaire texts based on surface expressions. They used the Maximum Entropy method for the task.

In this paper we describe two methods for intention identification of an utterance. First, we explain a method based on a similarity measure. In this method we compute a similarity between an input utterance and utterances in a corpus (case examples). We compare several similarity measures. Next, we describe a method based on the Maximum Entropy (ME) method. In addition, we discuss features for the methods and the effect of context information.

In Section 2, we explain our task, corpus and intentions. We prepare a manual for the tagging process of the intentions and evaluate the reliability of the tagging based on the manual. In Section 3, we describe the two methods for the intention identification. In Section 4, we evaluate the performance of each method and conclude this paper in Section 5.

2 Intention and Corpus

In this section we explain the construction of a corpus for our method and intentions in our task. Furthermore, we evaluate the validity of a tagging process for the corpus.

Proceedings of the 10th Conference of the Pacific Association for Computational Linguistics, pages 255–262.

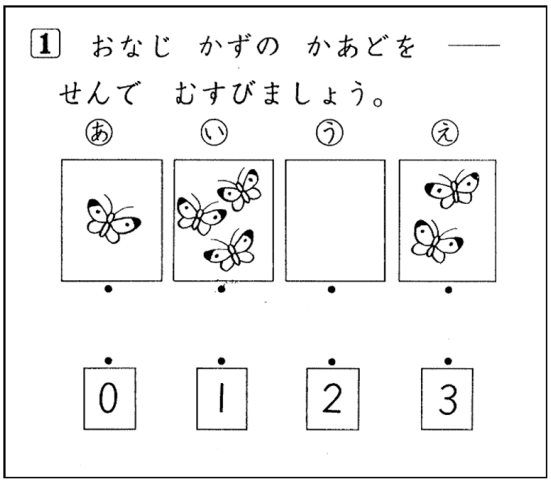
Table 1: The tagged corpus.

	# of turns	# of utterances
Teacher (T)	1390	1509
Student (S)	1449	1651
Total	2839	3160

2.1 Corpus

In this paper, we focus on a problem solving and knowledge acquisition system based on co-reference between drill texts and dialogue with a teacher (Endo and Kagawa, 1999; Shimada et al., 2007). We handled this task as a domain.

We collected a corpus consisting of simulated dialogues for this domain. The corpus consist of 81 dialogues with 27 test subjects for person-to-machine. The simulated dialogues for person-to-machine were collected by Wizard of OZ method (Fraser and Gilbert, 1991). We transcribed the dialogues manually, and then tagged fillers and ill-formed expressions, such as self-correction and hesitation, in them. Figure 1 shows an example of a drill text and a dialogue. The corpus contains 3160 utterances in 2839 turns (See Table 1). The turn denotes "S_i" and "T_i" in Figure 1.




```
S1: nanto onazi kazudesuka.
    (What is the number equal to?)
T1: kono waku no naka no chou wo kazoetekudasai.
    (Count the number of butterflies in this frame)
S2: kore wo kazoerunodesune.
    (May I count them?)
T2: soudesu.
    (yes)
S3: dokokara kaado wo erabunodesuka.
    (Where may I select from?)
T3: kokokara desu.
    (From here.)
S4: kono chou to kono kaado wo musubunodesune.
    (May I link card(1) and butterfly(1)?)
    .....
    ..... to be continued
```

Figure 1: An example of a drill text and a dialogue.

2.2 Intention

Various intention tags have been proposed in related studies (Araki et al., 1999; Irie et al., 2006; Searle, 1969; Walker and Passonneau, 2001). Walker and Passonneau (2001) have proposed a dialogue act tagging scheme for a travel domain. Their tags consisted of three dimension: (1) a speech act dimension, (2) a task-subtask dimension, and (3) a conversational domain dimension. Irie et al. (2006) have described a design of speech intention tags for an in-car spoken dialogue corpus. The tag set contained four layers: (1) Discourse act, (2) Action, (3) Object, and (4) Argument layers. These intention tags are specialized for a task. As a result, their tags support to determine an operation of a speech understanding system. However, most of the intention tags designed by them are domain-specific tags.

On the other hand, there are familiar speech-act labels and intentions, such as "request", "assert" and "question" (Araki et al., 1999; Searle, 1969). These intentions are more abstract than the domain-specific intentions. The purpose of our study in this paper is comparison of several methods for identifying the intention of an utterance. Therefore we use the traditional intention tags and expand them in this paper.

For the corpus we tag an intention label for each utterance. The number of intention labels is 12. Some intentions possess an attribute. The 5 intentions out of 12 possess attributes. The total number of intentions including attribute patterns is 18. Table 2 shows the intention tag set in our task.

We prepare a manual for the tagging process. The manual contains the explanation of the intention labels, examples of utterance-intention pairs and the guideline of the tagging. In this paper we tag one intention label to one utterance. Figure 2 shows an example of utterances and their intention tags.

2.3 Evaluation of the Corpus

We need to consider the reliability of the tagging process. For the evaluation the Kappa coefficient (κ) has been adopted as a standard in the dialogue processing (Cohen, 1960). The Kappa coefficient is computed as follows:

$$K = \frac{P(O) - P(E)}{1 - P(E)} \quad (1)$$

Table 2: Intention tags.

Label	Attribute	Definition
INFREQ (INFormation REQuest)	Howto	Request for how to solve a problem
	Request	Request for information to solve a problem
INFSUP (INFormation SUPplement)	Howto	Supplement for how to solve a problem
	Confirm	Supplement for information to solve a problem
AGREXP (AGReement EXPression)	Correct	Agreement for a confirmation request
	Response	Agreeable response
ACTREQ (ACTion REQuest)	Solve	Request for an action to solve a problem
	Inform	ACTREQ:solve with some hints
	Other	ACTREQ except above
CONREQ (CONfirmation REQuest)	Inform	Request for confirmation of information to solve a problem
	Correct	Request for confirmation: right or wrong
ACTREP (ACTion REPort)	——	Report for an action request
DISEXP (DISagreement EXPression)	——	Expression of disagreement
BEGDIA (BEGinning of DIAlogue)	——	Beginning of a dialogue
ENDDIA (END of DIAlogue)	——	End of a dialogue
REQRES (REQuest for REStating)	——	Request for restating an utterance; e.g. paraphrasing
ACTRES (ACTion REStating)	——	Action restating an utterance
OTHER	——	An insignificant utterance; e.g. filler only

Table 3: Kappa coefficient.

	Experience	NoExperience
$P(O)$	0.851	0.745
$P(E)$	0.027	0.020
Kappa coefficient	0.846	0.740
Average	0.793	

where $P(O)$ is the proportion of times the annotators agree and $P(E)$ is the proportion of times that we would expect the annotators to agree by chance. If the Kappa coefficient is close to 1, the degree of agreement is high.

We evaluated our tagging process with 10 test subjects. 5 persons out of 10 have tagging experience for other utterances in our task. For this evaluation we extracted 6 dialogues (210 utterances) from the corpus randomly. Table 3 shows the results of the evaluation. We obtained $\kappa = 0.793$ on average. In general $\kappa > 0.7$ indicates the substantial agreement of the tagging process (Araki et al., 1999). The result shows the high reliability of the tagging process with our manual.

3 Identification of Intention

In this section we explain two methods for intention identification: similarity measures and the maximum entropy method.

3.1 Similarity Measures

We compute a similarity between an input utterance and utterances in a corpus. We have already compared nine similarity measures: the inner product, some versions of the cosine measure, the Dice coefficient and the Jaccard coefficient (Iwashita et al., 2007) . In this paper we use three similarity measures that obtained high accuracy in the nine similarity measures. The similarities are the inner product (Inn), the Dice coefficient (Dice) and correspondence of morphemes (CorM). The correspondence of morphemes (CorM) is a similarity measure reported in related work (Irie et al., 2003). These similarities are computed as follows:

$$Inn(U_x, U_y) = \sum_{i=1}^{T} x_i \cdot y_i \qquad (2)$$

S_i: Dou sureba ii desuka? :: INFREQ:Howto
(How should I do?)

T_i: Mazu chou no kazu wo kazoete kudasai. :: ACTREQ:Solve
(First, count the butterflies.)
Soshite onazi kazu no kaado wo sagashite kudasai :: ACTREQ:Solve
(Then, search the card containing the same number.)

S_{i+1}: ..... to be continued.

Figure 2: An example of utterances and their intentions.

where U is an utterance. x_i and y_i are the value of a word i in U respectively. For the inner product, the value is a binary indicator, namely 1 (i exists in U) or 0 (otherwise). T is the number of vectors.

$$Dice(U_x, U_y) = \frac{2\sum_{i=1}^{T} x_i \cdot y_i}{\sum_{i=1}^{T} x_i^2 + \sum_{i=1}^{T} y_i^2} \quad (3)$$

For the Dice coefficient, we use word frequency for x_i and y_i in U_x and U_y.

$$CorM(U_x, U_y) = \frac{2M_{xy}}{M_x + M_y} \quad (4)$$

where M_x and M_y are the number of morphemes in U_x and U_y. M_{xy} denotes the number of morphemes matched between U_x and U_y. This computation is the binary vector version of the Dice coefficient.

The identification process with the similarity measures is as follows:

1. divide utterances into words by using the morphological analyzer ChaSen (Matsumoto et al., 1999)
2. construct a vector space for the identification process
3. compute the similarity between an input utterance (U_x) and each utterance in a corpus (U_y)
4. decide by a majority vote if some intentions possess the same maximum similarity.

3.2 Maximum Entropy

Maximum entropy modeling (ME) is one of the best techniques for natural language processing (Berger et al., 1996). The principle of the ME is expressed as follows:

$$P_\Lambda(c|d) = \frac{1}{Z_\Lambda(d)} \exp\left(\sum_i \lambda_{i,c} f_{i,c}(d,c)\right) \quad (5)$$

$$Z_\Lambda(d) = \sum_{d,c} \exp\left(\sum_i \lambda_{i,c} f_{i,c}(d,c)\right) \quad (6)$$

where $Z_\Lambda(d)$ is a normalization function. $\Lambda = \{\lambda_1, ..., \lambda_n\}$ are parameters for the model. These parameters denote weights and significance of each feature. The parameter values are a set that maximizes the entropy concerning the classifier. $f_{i,c}(d,c)$ is a feature function that is defined as follows:

$$f_{i,c}(d,c') = \begin{cases} 1 & \text{if } exist(d,i) > 0 \text{ and } c' = c \\ 0 & otherwise \end{cases} \quad (7)$$

where $exist(d,i)$ is a indicator function. The value is 1 in the case that a feature i exists in a document d.

In this paper we use Amis, which is a parameter estimator for maximum entropy models[1]. We estimate parameters by using the generalized iterative scaling algorithm.

The identification process with the ME is as follows:

1. divide utterances into words by using the morphological analyzer ChaSen
2. extract features from a corpus and then construct a training dataset
3. estimate parameters by using the ME method
4. output the intention that possesses the maximum probability computed from the estimated model

3.3 Features

We employ two types of features for the vector space model: (1) Bag-of-words (BOW) and (2)

[1] http://www-tsujii.is.s.u-tokyo.ac.jp/amis/index.html

Table 4: Accuracy.

	Similarity			ME
	Inn	Dice	CorM	
Q1	80.23	81.53	83.21	**84.98**
Q2	90.43	88.72	89.11	**92.32**
Q3	89.13	88.43	89.14	**92.42**
Ave	86.60	86.22	87.16	**89.91**

Table 5: Coverage.

	Similarity			ME
	Inn	Dice	CorM	
Q1	**88.74**	84.38	86.19	85.21
Q2	**96.01**	92.59	93.00	92.32
Q3	**93.96**	90.11	90.82	92.42
Ave	**92.92**	89.03	90.00	89.98

Dependency trees. We use all words as vectors for the BOW. For the dependency trees, we analyze utterances by using Japanese dependency analyzer Cabocha (Kudo and Matsumoto, 2002). We use frequent sub-trees for the vector space. We employ the FREQT algorithm (Asai et al., 2002) to extract them. The FREQT is an efficient pattern mining algorithm for discovering all frequent tree patterns from a large collection of labeled ordered trees. It is based on the rightmost expansion, a technique to grow a tree by attaching new nodes only on the rightmost branch of the tree.

4 Experiments

In this section, we evaluate the proposed method with our corpus. In this experiment, we evaluated the following points: (1) comparison of the similarity-based and the ME methods, (2) the coverage rates of each method, (3) the effectiveness of the sequence of utterances and (4) the effectiveness of dependency trees.

We evaluated this task with 81-fold cross validation, namely leave-one-out cross-validation. In other words, we used 1 dialogue as test data and 80 dialogues as case examples from 81 dialogues in our corpus. In this experiment we used the FREQT system implemented by T. Kudo[2].

First we compared the similarity measures and the ME-based methods. Table 4 shows the experimental result. The ME-based method outperformed the similarity-based methods. For the similarity-based methods, there was no significant difference in them although the accuracy of the related work (CorM) was a little better than those of the inner and the dice similarities.

However, the similarity-based methods often contained several intentions that possessed the same maximum similarity. We focused on the coverage rate of each method. The coverage was computed as follows:

$$Coverage = \frac{IncCorrect}{\mathrm{N}} \tag{8}$$

where N is the number of utterances in test data. *IncCorrect* is the number of utterances that contained the correct intention at least one. Table 5 shows the experimental result. Since the ME method usually output an unique intention as the 1st result, the difference between the accuracy and the coverage was slight. For this criterion, the inner product produced the best performance. This result denotes that the method with the inner product contains a possibility that it outperforms the ME method essentially.

Regarding all the methods, the accuracy of the problem Q1 was lower than those of Q2 and Q3. On our observation, this tendency depends on the difficulty of a problem. Q2 and Q3 were a simple problem, such as counting only. However Q1 needed to combine some processes[3]. To solve a difficult problem, a teacher and a student usually require enough interaction. Also, utterances in the dialogue often tend to be long and complex sentences. As a result, the accuracy of intention identification decreased.

In this paper we employed a majority vote for the final output of each similarity-based method. However, the frequencies of each intention tag in a corpus are not equable. For example, the tag <AGREXP:Response> in the corpus was 1/10 of the <AGREXP:Correct> tag. This denotes that the method probably outputs the <AGREXP:Correct> even if both of them exist in the intentions estimated by it. Therefore we introduced a normalization factor for the voting process.

$$v(int) = \frac{\#\text{ of } int \text{ in the output}}{\#\text{ of } int \text{ in the corpus}} \tag{9}$$

[2] http://cl.aist-nara.ac.jp/~taku-ku/software/freqt/

[3] Actually Figure 1 was the Q1 in this experiment. To solve this problem, a student needs at least 3 steps: counting butterflies, detecting the same number, and linking the cards.

Table 6: The use of a normalization factor for Inn.

	Majority	Eq. (9)
Q1	80.23	45.30
Q2	90.43	62.36
Q3	89.13	55.61
Ave	86.60	54.42

where *int* is an intention. Table 6 shows the experimental result. The normalization factor was not effective although it was effective for some intentions, such as <AGREXP:Response>.

One of the solutions for this problem, i.e. low frequency intention tags, is utilization of context information, such as tag sequence. Next, we evaluated the effectiveness of the sequence of utterances. Generally, the intention of an utterance depends on the intentions of the previous utterances. The history of intentions is one of the most important context in a dialogue. Matsubara et al. (2002) and Irie et al. (2003) have reported the effectiveness of context information, i.e. intention n-gram probability.

We applied the intention tag of the adjacent utterance in a corpus to our methods. First we extracted intention patterns that are the same as the previous intention of the input utterance. Next we computed the similarity between the input and the utterances extracted from a corpus. We computed a similarity measure (or a probability for ME) for all utterances in the corpus if no intention patterns exist in the corpus. Finally we decided by a majority vote if some intentions that possess the maximum similarity exist.

For the ME method, we generated a pair which consists of the intention of the current utterance and the intention of previous utterance. We regarded the pair as a new intention. For example, assume that the intentions of U_{i-1} and U_i are REQRES and ACTRES respectively. In this situation, we consider the intention of U_i to be REQRES-ACTRES. We computed the maximum entropy model by using these pair tags. In this experiment we substituted this method for the extraction of candidates, namely utilization of a tag sequence.

Table 7 shows the experimental result. In the table, "ON" denotes that we used the tag sequence for the method. In this experiment, we employed the inner product as the similarity measure. As a result, the utilization of tag sequence was ineffective for our dataset[4]. The accuracy decreased because utilizing a tag sequence led to the decrease of candidates for the similarity calculation[5]. One of the reasons is that our intention tags were simpler than that of the related work (Irie et al., 2003). Their intention tags consist of four hierarchized relations. Although the accuracy for intention identification including lower-level intention, namely the argument layer, increased by using context information, that for only the highest level intention did not improve. Our intention tags are similar to the highest level intention, namely the discourse act, in the related work. Therefore utilization of context information for this experiment did not lead to the improvement of the accuracy.

Table 7: The utilization of tag sequence.

	Inn		ME	
Use	OFF	ON	OFF	ON
Q1	**80.23**	80.12	**84.98**	82.51
Q2	**90.43**	88.16	**92.32**	90.33
Q3	**89.13**	89.04	92.42	**92.55**
Ave	**86.60**	85.77	**89.91**	88.47

Finally we evaluated the effectiveness of dependency trees. We use the inner product with binary vectors as the similarity measure. We compared two patterns: the minimum support and length of trees. The patterns are as follows:

Deps 1 the minimum support: 5, the pattern length: 2 to 5.

Deps 2 the minimum support: 20, the pattern length: 2 to 5.

The results are shown in Table 8. The result shows that using dependencies was ineffective. For several dialogues, the method with dependencies, however, outperformed that with the BOW only. Therefore we need the detail error analysis to obtain higher accuracy.

5 Discussion and Conclusions

In this paper we described a method for identification of the intention of an utterance. Our methods

[4]Needless to say, it was effective for some intentions, such as <AGREXP:Response>. However they did not lead to the improvement of accuracy because they were minorities in the test data

[5]In this experiment, the coverage rate also decreased. The coverage of the method with the inner product and context information was 88.67%. The original coverage was 92.92%. See Table 5.

Table 8: Accuracy with dependencies.

	BOW	Deps1	Deps2
Q1	80.23	80.31	**80.99**
Q2	**90.43**	89.29	89.53
Q3	**89.13**	88.45	88.12
Ave	**86.60**	86.21	86.02

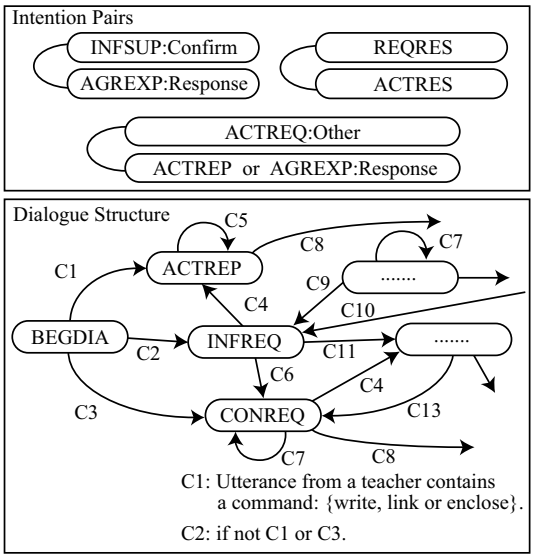


Figure 3: Intention pairs and A dialogue structure in our corpus (Part of).

computed a similarity (or a probability) between an input utterance and utterances in a corpus. In this experiment, the Maximum Entropy method outperformed similarity-based methods. However, the similarity-based method with the inner product sometimes contained the correct intention in the output. This denotes a possibility that the method with some adjustments might improve the accuracy. For the final output of similarity-based methods, we employed a majority vote. Although we used a normalization factor for the voting process, it led to the decrease of the accuracy. In the experiment, tag sequence and dependencies were not effective for the similarity calculation. However, they are effective for some intentions. Therefore we need to consider the usage of them.

One of the solutions to improve the performance is utilization of dialogue/discourse structures. Kato et al. (2005) have reported a dialogue annotation of an in-car speech corpus. They expressed the dialogue structure as a binary tree. In our task, intentions also contain a relation between them: a pair relation. We are developing a dialogue structure based on the analysis of our corpus. The dialogue structure is expressed by a transition network. Figure 3 shows a part of the relations and the dialogue structure The network contains the condition for the transition. We think that the relations and the structure are more useful than the context information that was described in Section 5 because they are global constraints in the dialogue. Effective utilization of relations between intention labels is one future work.

Table 9: Accuracy with a small corpus.

	Inn	Dice	CorM
Q1	76.47	74.79	76.38
Q2	85.10	85.87	86.02
Q3	84.67	84.78	84.13
Ave	82.08 (-4.52)	81.81 (-4.41)	82.18 (-4.98)

One of the approaches for the improvement of the accuracy is to incorporate high-level/abstract knowledge and heuristics. Matsubara et al. (2002) have used a word class for the similarity calculation. Also, in Japanese, clues of some intentions often appear in the end of the sentence. To weight words of the sentence-end might serve an important role in this task. Furthermore, we need to compare the method in this paper with other similarity measures and statistical techniques, such as decision tree learning (Irie et al., 2004) and Support Vector Machines (Vapnik, 1999). Integration of several similarity measures and machine learning methods is one of exciting approaches, such as the boosting algorithm (Freund and Schapier, 1996).

For our method, the size of the corpus is one of the most important problem. We evaluated the methods in the paper with a small dataset consisting of 27 dialogues. The result is shown in Table 9. The values in the parentheses in the table denote the difference of the accuracy from 81 dialogues (See Table 4). The accuracy decreased in the case that the corpus was small. Hence, in future studies, we need to develop a corpus tool kit for constructing a large corpus. Evaluation for other domains with the proposed method is also our future work.

References

M. Araki, T. Kumagai T. Ito, and M. Ishizaki. 1999. Proposal of a standard utterance-unit tag-

ging scheme (in Japanese). *Journal of JSAI*, 14(2):251–260.

T. Asai, S. Kawasoe K. Abe, H. Arimura, H. Sakamoto, and S. Arikawa. 2002. Efficient substructure discovery from large semi-structured data. In *Proceedings of the 2nd SIAM International Conference on Data Mining (SDM'02)*, pages 158–174.

A. L. Berger, S. A. Della Pietra, and V. J. Della Pietra. 1996. A maximum entropy approach to natural language processing. *Computational Linguistics*, 22(1):39–71.

J. Cohen. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1):37–46.

T. Endo and T. Kagawa. 1999. Cooperative understanding of utterances and gestures in a dialogue-based problem solving system. *Computational Intelligence*, 5(2):152–169.

N. Fraser and G. Gilbert. 1991. Simulating speech systems. *Computer Speech and Language*, 5(1):81–99.

Y. Freund and R. E. Schapier. 1996. Experiments with a new boosting algorithm. In *Proceedings of ICML*, pages 148–156.

R. Higashinaka, M. Nakano, and K. Aikawa. 2003. Corpus-based discourse understanding in spoken dialogue systems. In *Proceedings of ACL2003*, pages 240–247.

R. Higashinaka, K. Sudoh, and M. Nakano. 2005. Incorporating discourse features into confidence scoring of intention recognition results in spoken dialogue systems. In *Proceedings of ICASSP2005*, volume 1, pages 22–28.

H. Inui, M. Murata, M. Uchiyama, and H. Isahara. 2003. Classification of open-ended questionnaire texts based on surface expressions (in Japanese). *Journal of Natural Language Processing*, 10(2):19–42.

Y. Irie, S. Matsubara, N. Kawaguchi, Y. Yamaguchi, and Y. Inagaki. 2003. Speech intention understanding based on spoken dialogue corpus (in Japanese). In *SIG-SLUD-A301-03*, pages 7–12.

Y. Irie, S. Matsubara, N. Kawaguchi, Y. Yamaguchi, and Y. Inagaki. 2004. Speech intention understanding based on decision tree learning. In *Proceedings of the 8th International Conference on Spoken Language Processing*.

Y. Irie, S. Matsubara, N. Kawauguchi, Y. Yamaguchi, and Y. Inagaki. 2006. Layered speech-act annotation for spoken dialogue corpus. In *Proceedings of 5th International Conference on Language Resources and Evaluation*, pages 1584–1598.

K. Iwashita, K. Shimada, and T. Endo. 2007. Case-based estimation of speech intention (in Japanese). In *HINOKUNI symposium 2007 CD-ROM A-3-2*.

S. Kato, S. Matsubara, Y. Yamaguchi, and N. Kawaguchi. 2005. Dialogue structure annotation of in-car speech corpus based on speech-act tag. In *Proceedings of International Conference on Speech Databases and Assessment*, pages 159–163.

H. Kimura, M. Tokuhisa, K. Mera, K. Kai, and N. Okada. 1998. Comprehension of intentions and planning for responses in dialogue (in Japanese). In *IEICE Technical Report of Thought and Language, TL98-5*, pages 25–32.

T. Kudo and Y. Matsumoto. 2002. Japanese dependency analysis using cascaded chunking. In *Proceedings of the 6th Conference on Natural Language Learning 2002*, pages 63–69.

S. Matsubara, S. Kimura, N. Kawaguchi, Y. Yamaguchi, and Y. Inagaki. 2002. Example-based speech intention understanding and its application to in-car spoken dialogue system. In *Proceedings of the 17th International Conference on Computational Linguistics*, pages 633–639.

Y. Matsumoto, A. Kitauchi, T. Yamashita, Y. Hirano, H. Matsuda, , and M. Asahara. 1999. Japanese morphological analysis system chasen version 2.0 manual 2nd edition. Technical report, NAIST.

J. Searle. 1969. *Speech act*. Cambridge University Press.

K. Shimada, T. Endo, and S. Minewaki. 2007. Speech understanding based on keyword extraction and relation between words. *Computational Intelligence*, 23(1):45–60.

V. N. Vapnik. 1999. *Statistical Learning Theory*. Wiley.

M. Walker and R. Passonneau. 2001. Date: A dialogue act tagging scheme for evaluation of spoken dialogue systems. In *Proceedings of Human Language Technology Conference*, pages 66–73.