Criterion for Judging Request Intention in Response texts of Open-ended Questionnaires

INUI Hiroko
Communications Research Laboratory
Graduate School of Science and Technology Kobe University
hinui@crl.go.jp

UTIYAMA Masao
Communications Research Laboratory
mutiyama@crl.go.jp

ISAHARA Hitoshi
Communications Research Laboratory
Graduate School of Science and Technology Kobe University
isahara@crl.go.jp

Abstract

Our general research aim is to extract the actual intentions of persons when they respond to open-ended questionnaires. These intentions include the desire to make requests, complaints, expressions of resignation and so forth, but here we focus on extracting the intention to make a request. To do so, we first have to judge whether their responses contain the intent to make a request. Therefore, as a first step, we have developed a criterion for judging the existence of request intentions in responses. This criterion, which is based on paraphrasing, is described in detail in this paper. Our assumption is that a response with request intentions can be paraphrased into a typical request expression, e.g., "I would like to ...", while responses without request are not paraphrasable. The criterion is evaluated in terms of objectivity, reproducibility and effectiveness. Objectivity is demonstrated by showing that machine learning methods can learn the criterion from a set of intention-tagged data, while reproducibility, that the judgments of three annotators are reasonably consistent, and effectiveness, that judgments based not on the criterion but on intuition do not agree. This means the criterion is necessary to achieve reproducibility. These experiments indicate that the criterion can be used to judge the existence of request intentions in responses reliably.

1 Introduction

In every aspect of society, it is necessary for us to "know what the request is." This is because knowing what the request is plays an important role in allowing us to identify and solve problems to achieve improvements.

In recent years, the spread of electronic devices such as personal computers and the Internet has allowed us to save most requests in machine-readable texts. On the basis of these texts, research and development have been conducted "to know what the request is" as an element technology in natural language processing. For example, the research includes text mining (Nasukawa, 2001) and information extraction (Tateno, 2003) for customer claims and inquiries, development of an FAQ generation support system to a call center (Yanase et al., 2002; Matsuzawa, 2002), an FAQ navigation system using Q&A stored a call center (Matsui, 2002), and the development of requirement capturing methods for extracting requests made in meetings for software development (Doi, 2003). However, "to know what the request is" means to know the intention of various people in society such as residents, users, customers and patients, and it is inadequate to extract only request expressions expressed literally in texts. For this reason, previous works are not sufficient to understand intentions.

Against this background, (Inui et al., 1998; Inui et al., 2001; Inui and Isahara, 2002) have been studying how to extract and classify request intentions of respondents from responses of open-ended questionnaires (OEQs) which are accumulated requests. This paper describes the development of a criterion for judging request intentions and an evaluation of the criterion in terms of objectivity, reproducibility and effectiveness.

2 Development of the criterion for judging request intentions

2.1 Problems of an existing theory of modality

Response texts of OEQs are the focus of attention as data for text mining. Researchers have tried to extract various types of information from those texts (Lebart et al., 1998; Li and Yamanishi, 2001; Osumi and Lebart, 2000; Takahashi, 2000). However, they have mainly used only keywords (mostly nouns) as the basic units of extraction. If only the characteristic key words are analyzed with regard to sentences such as "Company A's beer tastes good," "Company A's beer does not seem to taste good," and "Company B's beer tastes better than company A's," the attention is directed toward "company A/company B/beer/tastes/good," and it is not possible to differentiate the meaning of the passages.

Because of this, as (Toyoda, 2002) points out, text mining in the future needs to treat modality, which often changes the meaning of the sentences completely. Two separate studies (Inui et al., 1998; Morohashi et al., 1998) have tried to process texts using words like auxiliary verbs and auxiliary verb equivalents as modality information. The modality information focused on in both studies, however, is grammatical expressions that have been accepted in a previous Japanese language study. Therefore, it is not possible to mechanically interpret requests and questions displayed by respondents, speakers and writers if they don't contain an auxiliary verb or an auxiliary verb equivalent.

In Japanese language syntax, modality is defined as the intention of the writer that is represented by grammatical expressions expressed grammatically (Nitta and Masuoka ed., 1989) and typically appears in the form of particles and auxiliary verbs in the sentence structure. Although previous text mining has focused on these expressions, modality does not always appear in the forms of grammatical expressions, and other expressions are more frequently used in real world texts. Thus, processing only those grammatical expressions listed so far is not sufficient for extracting intentions, and it is necessary to have a wide coverage of modality that expresses intentions.

2.2 Criterion to judge request intentions using paraphrasing

Surveyors try to know request intentions on the respondents through questionnaires, and respondents try to convey their request intentions to surveyors by responding to questionnaires. Therefore, it is important to establish a method that can extract the request intentions of the respondents based on the expressions given in the response texts. In this section, we propose a criterion to judge the existence of request intentions.

First, we will analyze the request expressions deductively. Native Japanese speakers can recognize expressions such as *te-hoshii* (would like you to), *te-moraitai* (would like you to), *te-kudasai* (please do) and *te-kure* (do) as request. These are linguistically called direct request expressions

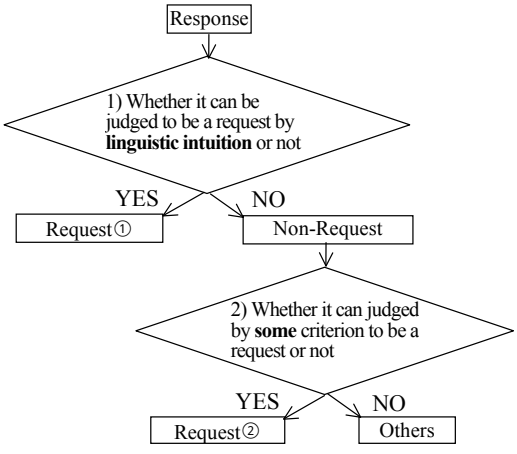

Fig. 1 Layers to judge expressions of requests

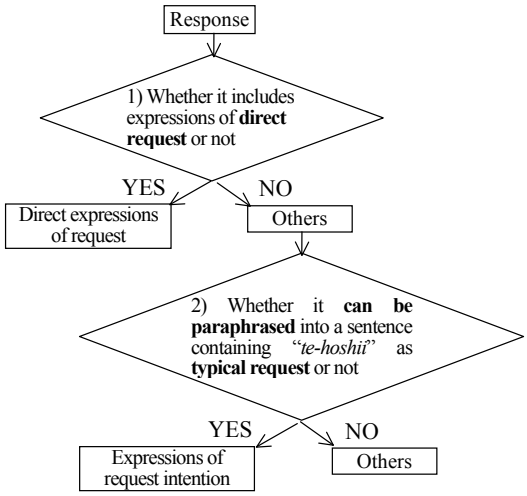

Fig. 2 Criterion to judge request intentions

(NIJLA, 1960) and able to indicate request intentions. Especially, *te-hoshii* is a typical request expression.

In other words, these direct request expressions are a clue to understand that there is a request intended. This recognition process is equivalent to the first judgment in Fig.1, that is, "whether a response can be judged to be a request by linguistic intuition or not." We regarded this as the first level criterion to judge request intentions. It corresponds to the first level in Fig.2, the intent of which is equal to judge whether the response includes a direct request expression or not.

Second, we consider the case that a response does not contain a direct request expression. In this case, non-requests in Fig.1 may be judged as requests. For example, based on the relation with surveyors, respondents and the situation, "Guardrails should be built along sidewalks of heavily congested roads" and "Building eco-friendly roads is important" can be interpreted as "We want guardrails along the sidewalks" and "We want you to think about the environment." However, the interpretation is due to "some" implicit criterion as shown in the second judgment in Fig. 1. As the implicit criterion depends on the judges, it is possible that the judgments differ[1]. This means that the results of the judgment, namely request ② in Fig. 1, are not re-created consistently. Therefore, the second judge in Fig.1 is not reproducible.

Consequently we attempted to manifest the implicit criterion as an explicit criterion to judge the existence of request intentions. This manifestation is the criterion "whether a response can be paraphrased into a sentence containing *te-hoshii* as a typical request expression or not" as the second judge in Fig. 2. As this criterion is explicit, the judgment of the criterion does not depend on the judges and agree consistently. Therefore, the second judge in Fig.2, namely the proposed criterion is reproducible and the results of the judgment, namely the expression of request intentions in Fig.2 is re-created consistently[2].

As mentioned above, we propose a criterion for judging request intentions by paraphrasing a response sentence into a typical request sentence contained *te-hoshii*. In Section 3, we evaluate the proposed criterion by a single judge analytically and objectively. In Section 4, we evaluate the results of experiments conducted by different judges from the viewpoint of reproducibility and effectiveness. These evaluations enable to demonstrate that the criterion, namely paraphrasing is an important method to determine the intentions independent of variety of surface expressions and differences among individual judgments.

3 Evaluation by a single judge

3.1 Analysis of response texts

Using the proposed criterion described in Section 2.2, we analyzed and classified response sentences manually according to two considerations: (1) if they include direct request expressions such as *te-hoshii* and *te-moraitai*; and (2) if it is possible to paraphrase them into a sentence ending with *te-hoshii*. To make the judgment for (1), we used request expressions listed by (Morita and Matsui, 1989).

	Expressions of direct requests	Paraphrase	Out of 3000 sentences
①	Included	Possible	547
②	Included	Not possible	3
③	Not included	Possible	1190
④	Not included	Not possible	1252

Table 1 Results of applying criterion for judging request intentions[3]

The analysis data are part of the response texts of OEQs carried out to make the best use of the opinions of the citizens in future road planning (Voice report, 1996). The original OEQ corpus contains a total of 35,674 respondents and 113,316 opinions. The analysis data comprised 3,000 sentences sampled at random after separating the plural sentences contained in the response text into single sentences. The criterion in Section 2.2 was used and the results are shown in Table 1.

Line ① in Table 1 includes sentences with direct request expressions such as *te-hoshii*, *te-kudasai* and *te-kure*. All of these could be paraphrased into *te-hoshii* and accounted for about 20% of the 3,000 sentences. Line ② includes direct request expressions that could not be paraphrased because they were used in quotations. These examples are exceptional. Expressions in line ③ correspond to

[1] This is demonstrated by the results of the experiment described in Section 4.2.

[2] This reproducibility is described in detail in Section 4.1.

[3] Eight sentences were excluded from Table 1 because they were ambiguous out of contexts.

Type of POS	Types of form of expression	Example Sentence
End-form in verbs and adjectives	-見やすくする(make…to do) /-取り締まる (control) etc.	緑地帯を多くし、標識をわかりやすく見やすくする。(Increase greenbelt and make it easier to see signposts)
Used as noun	-確保(secure)/-整備(equipment) etc.	駐車場の確保。(Secure car parks)
Predicates abbreviated	-を etc.	老人や子供や障害者の立場での道づくりを。(Road building from the standpoint of the elderly, children, and the disabled)
Verbs and adjectives of expectation and desire	-を求める (seek) /-に期待する (expect) / -願いたい (desire) /-が望ましい (is desirable) /-が望まれる (is desired)/-を望む (desire) /-を要望する(request) etc.	障害者、老人、子供、立場の弱い者が優先して通れる道が望まれる。(Roads and streets that give priority to the disabled, the elderly, children, and the weak are desirable)
Nouns for judging value	<attribute: *emergency*> -が急務である (matter of urgency) /-が最優先だ(first priority) /-が先決だと思う(think that the first thing to do) etc.	地方に高速道を建設するのもいいけど、渋滞箇所を整備していくことが先決ではないか。(It is all right to build expressways in provincial areas, but why can't improving congested places come first?)
	<attribute:*importance*> -が重要だ (is important) /-も大事な問題だと思う(think that it is also an important matter) /-が大切だ (is important) /-が大切だろう(should be important) /-が理想 (that is ideal) etc.	停車のマナーの徹底も大事な問題だと思います。(I think that the important matter is to make the manner of stopping vehicles thorough)
	<attribute: *necessity*> -ことも必要である (it may also be necessary) /-の必要を感じる (feel the necessity for) /-が不可欠だ(is indispensable) etc.	道づくりには、地権者の協力が不可欠です。(Cooperation of landowners is indispensable in road building)

Table 2 Expressions of requests and intention obtained by using the criterion for judging request intentions

expressions of request intentions in Fig.2 in Section 2.2. These expressions are shown in Table 2. Line ④ includes non-request expressions.

Table 2 shows various forms of expressions based on parts of speech (POS), i.e., verbs, nouns and adjectives, that have not been considered acceptable as modality expressions, even though they are paraphrasable by *te-hoshii*, and thus they are request expressions. As described in Section 2.1, several studies have been made on modality in terms of particles, auxiliary verbs, and auxiliary verb equivalents. However, little attention has been given to other POS in this regard. This is because modality expressions have been primarily connected with the grammatical elements such as auxiliary verbs in syntax. However, Table 2, which lists expressions of request intentions, shows that verbs, nouns and adjectives are actually also important elements that express modality.

Previous works that aim to extract requests have used pattern matching methods, and patterns that mainly consist of the direct request expressions corresponding to ① in Table 1. However, the results of manual analysis for paraphrasability shown in Table 2 indicate that using the proposed criterion enables many expressions of request intentions to be extracted from responses. In addition, we found a tendency for the number of expressions of request intentions direct request expressions, as shown in Table 1. In this section, we have provided explanation for the coverage of the criterion by analyzing response texts.

3.2 Evaluation of objectivity through machine learning methods

This section shows that the possibility of paraphrasing is learnable by machine learning methods. The data for the machine learning methods were tagged by the expert that analyzed the data in Table 1. Our assumption is that if machine learning methods can learn the paraphrasability from the data, then the data are said to have been tagged consistently enough to be mechanically learnable. This indicates that the criterion proposed in Section 2 is objectively applicable to tag data.

Machine learning methods

We use two machine learning methods in this section. They are maximum entropy method (ME) (Beger et al. 96) and support vector machine (SVM) (cristianini00)[4], both of which have been shown to be quite effective in natural language processing.

The task of a machine learning method is to make a classifier that can decide whether a response is paraphrasable by *te-hoshii* or not. A response *X* is tagged *possible* if it is paraphrasable

[4] We used maxent (http://www.crl.go.jp/jt/a132/members/mutiyama/software.html) for ME learning and TinySVM(http://cl.aist-nara.ac.jp/~taku-ku/software/TinySVM/) for SVM learning.

and *impossible* if not. X is represented by a feature vector $x = [x_1, x_2, \ldots\ldots, x_l]$where

$$x_i = \begin{cases} 1 & \text{if } X \text{ has feature } i \\ 0 & \text{otherwise} \end{cases}$$

Given training data, a machine learning method produces a classifier that outputs *possible* or *impossible* according to a given feature vector. We omit the details of ME and SVM. Readers are referred to the above references.

We will compare three sets of features, F_1, F_2 and F_3, in the experiments below. F_1 consists of word 1-grams, F_2, 1-grams and 2-grams, and F_3, word 1-grams, 2-grams and 3-grams. For example, let X be a response consisting of a word sequence[5] w_1, w_1,….., w_m where w_1 = 〈b〉 and w_m = 〈e〉 are special symbols representing the beginning and the ending of a response. Let S_1 be the set of 1-grams in X $\{w_i \mid 2 \leqq i \leqq m\text{-}1\}$, S_2, 2-grams in X $\{ w_i w_{i+1} \mid 1 \leqq i \leqq m\text{-}1\}$ and S_3, 3-grams in X $\{ w_i w_{i+1} w_{i+2} \mid 1 \leqq i \leqq m\text{-}2\}$. The F_1, F_2 and F_3 features contained in X are S_1, $S_1 \cup S_2$, and $S_1 \cup S_2 \cup S_3$, respectively.

Experiments

The data used for the experiments consisted of 3,001 responses[6]. The numbers of the responses tagged *possible* and i*mpossible* were 1,944 and 1,057, respectively. We used 10-fold cross validation to evaluate the accuracies of ME and SVM[7]. For each iteration in the cross validation, 8/10 of the data was used for training, 1/10, for parameter adjustment, and 1/10, for testing. The precision, P_i, for iteration i is

$$P_i = \frac{\text{number of correctly tagged answers}}{\text{total number of answers in the test data}}$$

We define P as the mean of the precisions for each iterations, i.e., $P = \Sigma_i P_i/10$. We henceforth call P precision. The precisions of ME and SVM are in Table 3, together with a baseline precision 0.648 (=1944/3001), which was obtained by tagging all the responses *possible*. In the table, the figures in columns "ME" and "SVM" are the precisions of ME and SVM. Line F_i (i=1,2,3) indicates that the precisions in that line were obtained by using F_i as a feature set. We use one-sided Welch tests to measure the differences between precisions and say "statistically significant" or simply "significant" when the differences were statistically significant at 1% level.

Table 3 indicates that both ME and SVM outperform the baseline by a large margin. The differences were, of course, statistically significant. Therefore, we can conclude that these methods are quite effective in this task.

	ME	SVM	Baseline
F_1	0.892	0.887	0.648
F_2	0.912	0.909	0.648
F_3	0.913	0.915	0.648

Table 3 Precision of ME and SVM

This table also indicates that ME and SVM are comparable in precision. The differences of precision were not statistically significant. We next compared the highest precisions in lines F_1, F_2, and F_3. F_1 was significantly outperformed by both F_2 and F_3, but there was not a significant difference between F_2 and F_3. Consequently, we can use either ME or SVM as a machine learning method and F_2 or F_3 as a feature set.

Table 3 demonstrates that we can expect about 91% precision in deciding the paraphrasability by using either ME or SVM. This is a reasonably high precision. Therefore, we can conclude that the criterion proposed in Section 2.2 is sufficiently objective and stable.

4 Evaluation by different judges

In Section 3, we described the manual analytical evaluation by a single judge and the objective evaluation by machine learning that uses a corpus prepared based on the analytical evaluation. Section 4 refers to experiments carried out by multiple different judges.

4.1 Evaluation of reproducibility: judgment of paraphrasing by multiple judges

The subjects of this experiment were three male native speakers of Japanese in their twenties who were engineering majors. The experiment was carried out using a total of 24,000 random sentences from the OEQ corpus described in Section 3.1 by applying the criterion proposed in Section 2.2. If a response text included plural sentences, they were separated into single sentences as mentioned in Section 3.1. Of the

[5] We used ChaSen (http://chasen.aist-nara.ac.jp/) to segment an answer into a word sequence.

[6] This data was different from the response text analyzed in Section 3.1.

[7] We used the polynomial kernel for SVM. We tried degrees 1 and 2 d=1,2. Since d=1 outperformed d=2, the results of d=1 are in Table 3

24,000 sentences, the three subjects A, B and C were each given 8,000 of them. However, the pairs A and B, B and C, and A and C were each given 4,000 common sentences, so that a variation of sentence totaled 12,000.

As shown in Table 1 in Section 3.1, direct request expressions can be paraphrased with *te-hoshii*, therefore, we deal only with the judgment of the second level in Fig.2, namely the paraphrasing into *te-hoshii*. For the evaluation, we prepared a set of work instructions for the subjects, part of which is shown below.

Work instructions

1) Not only the end expression but also case particles, case particle equivalents and those containing such expressions or expressions of connection are to be paraphrased.

2) If *te-hoshii* is to be changed to a negative request of *shite-hoshiku-nai* (do not want), place the word *negative* at the end.

3) Not only functional words but also content words, furthermore, word order may be changed in paraphrasing

#1 S(ource): 駐車場が少なすぎると思う (We think that there are not enough car parks.)

→ T(arget): 駐車場を増やしてほしい(We want car parks to be increased.)

The experimental results are given in Table 4, where P means *possible* to paraphrase and NP means *not possible*. KC is the kappa coefficient between subjects (Cohen 1960).

	B			
A	P	NP	Total	KC
P	2372	970	3342	0.48
NP	36	622	658	
Total	2408	1592	4000	
	C			
A	P	NP	Total	KC
P	3123	264	3387	0.61
NP	171	442	613	
Total	3294	706	4000	
	C			
B	P	NP	Total	KC
P	2119	50	2169	0.49
NP	934	897	1831	
Total	3053	947	4000	

Table 4 Results of paraphrasability using the criterion

Generally, the closer the kappa coefficient is to 1, the higher the degree of agreement is obtained. There is a complete agreement when it is 1. In general, the ranges [0.81-1.00], [0.61-0.80], [0.41-0.60], [0.21-0.40] and [0.00-0.20] correspond to full, practical, medium, low, and no agreement, respectively.

Therefore, as Table 4 indicates, the results of the judging and the paraphrasing using the criterion by the three subjects showed that there was substantial agreement between subject A and C, and medium agreement between A and B, and B and C.

These results indicate that the method based on the criterion, whether used by a single judge or by different judges(=subjects) for analysis and experiment, enables requests and non-requests to be distinguished. Therefore, we can conclude that using the criterion enables even untrained people to reproduce the extraction of requests.

Sentences such as #2 and #3 below are examples of sentences that were agreed to be non-paraphrasable. These include expressions of intentions in which the current situation is accepted passively such as #2 ”しょうがない(I think that it cannot be helped),” or in which the current situation is actively accepted such as #3 ”素晴らしい(are wonderful)”. Furthermore, #4 is a sentence that begins with a clear statement of reason “理由は(the reason is).” This indicates that a motive for requests exists, and that a response formed by multiple sentences often composes request-motive adjacency in discourse structure.

Examples of sentences that could not be paraphrased:

#2 必要があれば料金の値上げもしょうがないと思う。(I think that it cannot be helped if rise in charges is necessary.)

#3 車いすの人でも、楽にあちこち一人で買い物や散歩ができる町、道路で素晴らしい (The town and roads are wonderful as even people in wheelchairs can do shopping by themselves here and there with ease and wander about.)

#4 理由は全体的な発展が望めないので。(The reason is that overall development cannot be hoped for.)

This analysis shows that paraphrasable sentences indicate requests and non-paraphrasable sentences indicate the acceptance of the current situations or the motives for requests.

4.2 Evaluation of effectiveness: judging intention without using the criterion

To evaluate whether the proposed criterion described in Section 2.2 is effective or not, we carried out an experiment to see if a response

shows requests or not without the criterion. The two subjects, D and E, who took part in this experiment were both native speakers of Japanese. Subject D was a male student in his twenties from the education department of a university, and subject E was a female student also in her twenties from the literature department of a university. They used the same data of 4,000 sentences that were used by the subjects B and C in Section 4.1. The subjects D and E did not consult with each other and carried out the work separately. We provided them with the following instructions before asking them to start the work.

- Each response sentence is context-free.
- Judge intuitively, and mark 1 if you think the sentence shows a request, and mark 0 if you do not .
- Make sure to mark either 1 or 0.

The results of the experiment are given in Table 5, where 1 and 0 in the right table correspond to P and NP in Table 4. We show the data again because subjects B and C used the same data as subjects D and E. In Table 5, the kappa coefficient (KC), between D and E is lower than that between B and C. Moreover, it is the lowest among all those given in Tables 4 and 5. The KC of 0.17 means there is no agreement between D and E.

The results indicate the rate of agreement is higher for judgments made using the criterion than for subjective judgments. That is to say, this proves the effectiveness of the criterion.

	E				C		
D	1	0	Total	B	P	NP	Total
1	562	1880	2442	P	2119	50	2169
0	39	1517	1556	NP	934	897	1831
total	601	3397	3998	total	3053	947	4000
KC for D&E		0.17		KC for B&C		0.49	

Table 5 Results for experiment for effectiveness

4.3 Examination of evaluation results

We examine here mainly the cases in which no agreement was obtained with respect to paraphrasing in the experiment described in Section 4.1. Table 4 shows the cases where disagreement was considerable. The results for these cases, shown in Table 6, indicate that disagreement is obtained when the sentences are paraphrased into the forms including clauses of cause and reason indicated by "*node*"(because) as #5. The clause is underlined in the target sentence in #5.

Examinees		A B	A C	B C
No. of paraphrase includes *node*		648	224	89
subject	A	645	194	---
	B	3	---	3
	C	---	30	86
No. of disagreed paraphrasing		1006	435	984
Rates of *node* in disagreements (%)		64.4	51.5	9.0

Table 6 Disagreed paraphrase including cause and reason clauses "*node*"

#5 S:狭い道路をますます狭くしている。(A narrow road is made even narrower)
T: 狭い道路をますます狭くしているので(*node*)、どうにかしてほしい。(Because the narrow road is made even narrower, I would like to see something done about it.)

The source sentence #5 is a statement showing the condition of the road being narrow. This statement can be seen as a motive for a request in the target sentence of #5. That is to say, the source sentence #5 itself shows not the content of a request but the "motive for request." The three subjects disagreed in their judgments on whether or not the "motive for request" sentence was paraphrasable as shown in the bottom line of Table 6. As the table indicates, disagreement rates of 64.4%, 51.5%, and 9.0% were obtained between A and B, A and C, and B and C. The reason for these high disagreement rates was that we did not give clear directions in the work instructions. The sentences which the paraphrasing includes "*node"* are not requests and should not be extracted. This means these sentences should have been considered to be non-paraphrasable.

On the other hand, with regard to "motive for request" sentences, there was an example #1 in Section 4.1 in which the work instructions requested the subjects to paraphrase such a sentence. That is, the work instructions suggested that the source sentence #1 "I think that we do not have enough car parks" is a motive for the request "I want car parks to be increased." This kind of inadequate instruction led to instability in the work done and might have increased the disagreement rates obtained in the judgment.

However, according to the data prepared by the expert referred to in Section 3.2, "motive for request" sentences cannot be paraphrased into *te-hoshii*, and machine learning has confirmed that the data are objective. Therefore, it can be considered that the work of removing "motive for requests" sentences can be done stably. This means

that if the work instructions give clear directions like "if you are able to add *node* at the end of a sentence, that sentence should be regarded not as a content of request, but a motive of request," then the rate of agreement may be improved.

5 Conclusion

We have developed a criterion for judging request intentions. We evaluated this criterion from three points of view. The first evaluation was to analyze the data applying the criterion by a single judge. From this analysis, it was found that this criterion makes it possible to extract requests and that the coverage can be guaranteed compared with previous studies. Moreover, a corpus was prepared based on the analysis and was used for a machine learning experiment. From this experiment results, we confirmed the criterion using a paraphrase was objective.

Furthermore, by different judges, the second evaluation was made from the experiment conducted by three subjects. The rate of agreement for the paraphrasability was high, which indicated that the results of requests extraction were re-created using the criterion. This proves the reproducibility of the criterion.

In the third experiment, two subjects judged the sentences without using the criterion to see whether or not there was a request in each response sentence. A comparison of the results of the second and the third experiments showed that a higher rate of agreement was obtained with the method using the criterion. This confirmed the effectiveness of the criterion.

In future work, we will analyze "motives for request" sentences found from the examinations, and prepare a criterion for distinguishing between request motives and the contents of request intentions.

References

Adam L. Berger, Stephen A. Della Pietra, and Vincent J. Della Pietra. 1996. A maximum entropy approach to natural language processing. *Computaional Linguistics*, Vol.22, No.1, pp39-71.

Jacob Cohen. 1960. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*. 20, 37-46.

Nello Cristianini and John Shawe-Taylor. 2000. *An Introduction to Support Vector Machines*. Cambridge University Press.

Kouichi Doi, Naoyuki Horai, Isamu Watanabe, Yoshinori Katayama and Masayuki. Sonobe. 2003. User-oriented Requirements Capturing Method in Analyzing Requirements Capturing Meeting. *Transactions of IPSJ*, vol.44 No.1, pp48-58.

The Committee for Roads in the 21st Century Basic Policy Board, Road Council. 1996. *Voice Report*.

Hiroko Inui, Kiyotaka Ucihmoto and Hitoshi Isahara. 1998. Classification of Open-Ended Questionnaires based on Analysis of Modality. *Proceedings of the 4th Annual Meeting of the ANLP*, pp540-543.

Hiroko Inui, Masaki Murata, Kiyotaka Uchimoto and Hitoshi Isahara. 2001. Classification of Open-Ended Questionnaires based on Surface Information in Sentence Structure. *Proceedings of the 6th NLPRS2001*, pp315-322.

Hiroko Inui and Hitoshi Isahara. 2002. Proposition for "Extended Modality" –Extraction of Intention in Open-ended response texts-. *Technical Report of EICE*, Vol.102 No.414, NLC2002-43, pp31-36.

Ludovic Lebart, Andre Salem and Lisette Berry. 1998. *Exploring Textual Data*, Kluwer Academic Publishers, 14-20.

Hang Li and Kenji Yamanishi. 2001. Mining from Open Answers in Questionnaire Data Using Statistical Learning Techniques. Proceedings of the 4 IBIS2001. pp129-134.

Kunio Matsui and Hozumi Tanaka. 2002. The Navigation to the Stored Q&A data using Simple Questions. *Technical Report of IEICE*, Vol.102 No.414, NLC2002-40, pp13-18.

Hirofumi Matsuzawa. 2002. FAQ Generation Support System Using Structured Association Pattern Mining and Natural Language Processing. *Proceedings of the FIT2002*, pp69-70.

Yoshiyuki Morita and Masae Matsuki. 1989. *Expression Pattarn of Japanese*, ALC

Masayuki Morohashi, Tetsuya Nasukawa and Touru Nagano. 1998. Text Mining: Knowledge Acquisition from enormous text data – recognition of intention -. *Proceedings of the 57th Annual Meeting of IPSJ*

Tetsuya Nasukawa. 2001. Text Mining Application for Call Centers. *Journal of the Japanese Society for Artificial Intelligence*, Vol.16, No.2, pp219-225.

Noboru Ohsumi and Ludovic Lebart. 2000. Analyzing Open-ended Questions: Some Experimental Results for Textual Data Analysis Based on InfoMiner. *Proceedings of the Institute of Statistical Mathematics*. Vol.48, No.2, pp339-376

The National Institute for Japanese Language. 1960. *A research for making sentence patterns in colloquial Japanese. 1. On materials in conversation.* Shuei Publishers.

Yoshio Nitta and Takashi Masuoka. 1989. *Japanese Modality*. Kurosio Publishers.

Kazuko Takahashi. 2000. A supporting System for Cording of the answers from Open-Ended Question. *Sociological Theory and Methods*, vol.15, No.1. 149-164.

Masakazu Tateno. 2003. The Method to extract Textual "Kansei" Expression in the Custmer's Voice. *IPSJ SIG* Notes, NL-153-14, pp105-112.

Yuki Toyoda. 2002. Translation from Text Data to Numeric Data –Points for Attention in Text Mining Preparatory Processing as Seen from the Analyst's. *Journal of the Japanese Society for Artificial Intelligence*, Vol.17 No.6. pp738-743.

Takashi Yanase, Satoko Marumoto, Isao Nanba and Ryo Ochitani. 2002. Parsing Question Texts Using the Predicate Expressions of the Sentence End. *Proceedings of the 8th Annual Meeting of the Association for NLP*, pp647-650.