PAPER Special Section on Corpus-Based Speech Technologies

Robust Dependency Parsing of Spontaneous Japanese Spoken Language

Tomohiro OHNO[†a)], ***Student Member***, **Shigeki MATSUBARA**[††], **Nobuo KAWAGUCHI**[††], ***Members***, ***and*** **Yasuyoshi INAGAKI**[†††], ***Fellow***

SUMMARY Spontaneously spoken Japanese includes a lot of grammatically ill-formed linguistic phenomena such as fillers, hesitations, inversions, and so on, which do not appear in written language. This paper proposes a novel method of robust dependency parsing using a large-scale spoken language corpus, and evaluates the availability and robustness of the method using spontaneously spoken dialogue sentences. By utilizing stochastic information about the appearance of ill-formed phenomena, the method can robustly parse spoken Japanese including fillers, inversions, or dependencies over utterance units. Experimental results reveal that the parsing accuracy reached 87.0%, and we confirmed that it is effective to utilize the location information of a bunsetsu, and the distance information between bunsetsus as stochastic information.

key words: *dependency parsing, stochastic parsing, Japanese speech, linguistic phenomena, syntactically annotated corpus*

1. Introduction

With the recent advances in continuous speech recognition technology, a considerable number of studies have been conducted on spoken dialogue systems. For the purpose of smooth interaction with the user, it is necessary for the system to understand spontaneous speech. Since spontaneously spoken language includes a lot of grammatically ill-formed linguistic phenomena such as fillers, hesitations and self-repairs, a technique for robust parsing is thus urgently required. Because the inflectional languages such as English have relatively strong grammatical constraints, it might be feasible to parse the spoken languages in a grammar-oriented fashion (for example, [1], [7]). On the other hand, such the approach are not necessarily suited to parsing Japanese spoken language because of its contrary linguistic features.

This paper describes the characteristic features of Japanese spoken language on the basis of investigating a large-scale spoken dialogue corpus from the viewpoint of dependency, and moreover, proposes a method of dependency parsing by taking such features into account. The conventional method of dependency parsing generally assumes the following three syntactic constraints [13]:

1. No dependency is directed from right to left.
2. Dependencies do not cross each other.
3. Each *bunsetsu**, except the last one, depends on only one bunsetsu.

As far as we have investigated the corpus, however, many spoken utterances do not satisfy these constraints because of inversion phenomena, bunsetsus that do not depend on any bunsetsu, and so on. Therefore, our parsing method relaxes the first and third of the above three constraints, that is, it permits the dependency directed from right to left and a bunsetsu that does not depend on any other bunsetsu.

The method acquires in advance the probabilities of dependencies from a spoken dialogue corpus tagged with dependency structures, and provides the most plausible dependency structure for each utterance on the basis of those probabilities. Several techniques for dependency parsing based on stochastic approaches have been proposed so far. Collins [3] and Fujio and Matsumoto [5] have used the probability based on the frequency of cooccurrence between two bunsetsus for dependency parsing. Ratnaparkhi [18], Uchimoto et al. [20] and Charniak [2] have applied a maximum entropy method to dependency or syntactic structure analysis. Furthermore, Haruno et al. [6] and Kudo and Matsumoto [11] have proposed a technique for learning the dependency probability model based on decision trees and SVMs respectively.

However, since these techniques are for written language, it remains unclear as to whether they are proper models for spoken language. As a technique for stochastic parsing of spoken language, Den has suggested a new idea for detecting and parsing self-repaired expressions; however, the phenomena with which the framework can cope are restricted [4].

On the other hand, our method provides the most plausible dependency structures for spontaneous speech by utilizing stochastic information. In order to evaluate the effectiveness of our method, we have conducted an experiment on dependency parsing. In the experiment, all drivers' utter-

Manuscript received July 1, 2004.
Manuscript revised September 27, 2004.

[†]The author is with the Graduate School of Information Science, Nagoya University, Nagoya-shi, 464–8601 Japan.
[††]The authors are with the Information Technology Center/CIAIR, Nagoya University, Nagoya-shi, 464–8601 Japan.
[†††]The author is with the Faculty of Information Science and Technology, Aichi Prefectural University, Aichi-ken, 480–1198 Japan.
a) E-mail: ohno@el.itc.nagoya-u.ac.jp

DOI: 10.1093/ietisy/e88–d.3.545

*A *bunsetsu* is one of the linguistic units in Japanese, and roughly corresponds to a basic phrase in English. A bunsetsu consists of one independent word and more than zero ancillary words. A *dependency* is a modification relation that a *dependent bunsetsu* depends on a *head bunsetsu*. That is, a dependent bunsetsu and a head bunsetsu work as a modifier and a modifyee, respectively.

ances in 81 spoken dialogues of CIAIR in-car speech dialogue corpus [8]–[10] have been used. The experimental result has shown our method to be available for robust parsing of spontaneously spoken language. In particular, we report how effectively the method can parse bunsetsus that have no head bunsetsu, dependencies directed from right to left, and dependencies over utterance units.

This paper is organized as follows: The next section explains linguistic analysis of spontaneous Japanese speech. Section 3 describes robust dependency parsing of spoken language. Section 4 presents the parsing experiment. The discussion about the robustness of our method is reported in Sect. 5.

2. Linguistic Analysis of Spontaneous Speech

We have investigated spontaneously spoken utterances in an in-car speech dialogue corpus, which was constructed at the Center for Integrated Acoustic Information Research (CIAIR), Nagoya University [8]–[10]. The corpus contains speech dialogue between drivers and navigators and their transcripts.

2.1 CIAIR In-Car Speech Dialogue Corpus

The Center for Integrated Acoustic Information Research (CIAIR), Nagoya University has been collecting large-scale in-car speech dialogues [8]–[10]. During the project CIAIR members have developed a Data Collection Vehicle (DCV), which is shown in Fig. 1, and collected a total of about 400 GB of data by recording three sessions of spoken dialogue by 800 drivers in drives of about 60 minutes each.

A spontaneously spoken language corpus has been

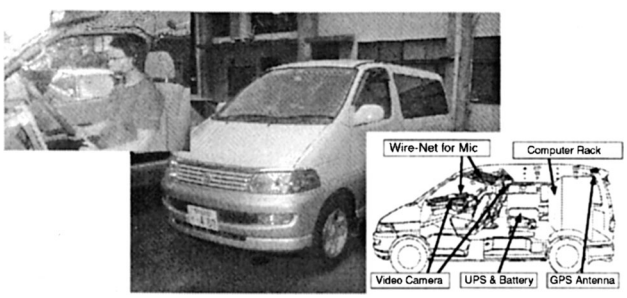

Fig. 1 The data collection vehicle (DCV).

0003 – 00:09:382-00:13:652 F:D:I:I:		
今日	[today]	& キョー
朝	[morning]	& アサ
パン	[bread]	& パン
食べて	[ate]	& タベテ
お昼は	[lunchtime]	& オヒルワ
おそばを	[soba]	& オソバオ
食べたんですよ<H><SB>	[ate]	& タベタンデスヨ<H><SB>
0004 – 00:13:995-00:17:328 F:D:I:I:		
今晩は	[tonight]	& コンバンワ
何	[what]	& ナニ
食べよっかな<SB>	[want to eat]	& タベヨッカナ<SB>

Fig. 2 Transcript of in-car dialogue speech.

constructed by transcribing the collected speech data into ASCII text files by hand in accordance with the rule of the corpus of spoken Japanese (CSJ) [14]. The corpus is composed of in-car dialogues between drivers and navigators about shop retrieval, driving directions, and so on. An example of a transcript is shown in Fig. 2. For advanced analysis, discourse tags are assigned to fillers, hesitations, slips, and so on†. Furthermore, each speech is segmented into utterance units by a pause, and their exact start and end times are provided.

We have investigated all drivers' utterance units from 195 dialogues. The number per utterance unit of fillers, hesitations and slips, is 0.34, 0.07, 0,04, respectively. The fact that the frequencies are not less than those of usual human-human conversations suggests the in-car speech of the corpus to be spontaneous.

2.2 Dependency Structure of Spoken Language

To characterize spontaneous dialogue speeches from the viewpoint of dependency, we have constructed a syntactically annotated spoken language corpus by providing morphological and syntactic information for each of the driver's utterances in CIAIR in-car speech dialogue corpus [16]. The morphological and syntactic information were provided for the corpus.

Boundaries between words, pronunciation, basic form, part-of-speech, conjugation type and conjugated form of each word as the morphological information, and boundaries and dependencies between bunsetsus as the syntactic information were provided for the corpus.

Here, the specification of the parts-of-speech is in accordance with that of IPA parts-of-speech in a morphological analyzer called ChaSen [15], the rules of the bunsetsu segmentation with those of CSJ [14], and the dependency grammar with that of the Kyoto Corpus [12]. We have provided the following criteria for the linguistic phenomena peculiar to spoken language:

- There is no bunsetsu on which fillers and hesitations depend; they form dependency structures independently.
- A bunsetsu whose head bunsetsu is omitted does not depend on any bunsetsu.
- The specification for parts-of-speech has been provided for phrases peculiar to spoken language by adding lexical entries to the dictionary.
- We define one conversational turn as a unit of dependency parsing. The dependencies might be over two utterance units, but rarely over two conversational turns.

Figure 3 shows an example of spoken Japanese sentences annotated by a dependency structure. It illustrates a sequence of dependency relations, each of which consists of

†We assume that a speech recognizer can distinctively output disfluencies such as fillers and hesitations. Some methods for realize it have been proposed so far [17], [19]. In this paper, the part-of-speech of the disfluencies is treated as none.

```
((1 ((きょう kyo きょう noun 副詞可能 none none)))[today]
 -> (4 ((食べ tabe 食べる verb 自立 一段 連用形)
        (て te て particle 接続助詞 none none)))[ate] )

((2 ((朝 asa 朝 noun 副詞可能 none none)))[morning]
 -> (4 ((食べ tabe 食べる verb 自立 一段 連用形)
        (て te て particle 接続助詞 none none)))[ate] )

((3 ((パン pan パン noun 一般 none none)))[bread]
 -> (4 ((食べ tabe食べる verb 自立 一段 連用形)
        (て te て particle 接続助詞 none none)))[ate] )

((4 ((食べ tabe 食べる verb 自立 一段 連用形)
     (て te て particle 接続助詞 none none)))[ate]
 -> (7 ((食べ tabe 食べる verb 自立 一段 連用形)
        (た ta た auxiliary-verb none 特殊・タ 基本形)
        (ん n ん noun 非自立 none none)
        (です desu です auxiliary-verb none 特殊・デス 基本形)
        (よ yo よ particle 終助詞 none none)))[ate] )

((5 ((お昼 ohiru お昼 noun 副詞可能 none none)
     (は wa は particle 係助詞 none none)))[lunctime]
 -> (7 ((食べ tabe 食べる noun 自立 一段 連用形)
        (た ta た auxiliary-verb none 特殊・タ 基本形)
        (ん n ん noun 非自立 none none)
        (です desu です auxiliary-verb none 特殊・デス 基本形)
        (よ yo よ particle 終助詞 none none)))[ate] )

((6 ((お o お prefix 名詞接続 none none)
     (そば soba そば noun 一般 none none)
     (を o を particle 格助詞 none none)))[soba]
 -> (7 ((食べ tabe 食べる noun 自立 一段 連用形)
        (た ta た auxiliary-verb none 特殊・タ 基本形)
        (ん n ん noun 非自立 none none)
        (です desu です auxiliary-verb none 特殊・デス 基本形)
        (よ yo よ particle 終助詞 none none)))[ate] )

((7 ((食べ tabe 食べる verb 自立 一段 連用形)
     (た ta た auxiliary-verb none 特殊・タ 基本形)
     (ん n ん noun 非自立 none none)
     (です desu です auxiliary-verb none 特殊・デス 基本形)
     (よ yo よ particle 終助詞 none none))[ate] )
 -> (NO (なし)))
```

Fig. 3 Spoken Japanese sentence annotated by dependency structure.

Table 1 Corpus data for dependency analysis.

Utterance units	7,781
Conversational turns	6,078
Bunsetsus	24,250

a dependent bunsetsu and a head bunsetsu. Each bunsetsu is listed with its number and its constituent morphemes.

The outline of the corpus with dependency analyses is shown in Table 1. There exist 11,866 dependencies for 24,250 bunsetsus. The average number of dependencies per turn is 1.95, and is exceedingly lower than that of written language such as newspaper articles (about 10 dependencies). This does not necessarily mean that dependency parsing of spoken language is easier than that of written language. It is also necessary to specify the bunsetsu with no head bunsetsu because not every bunsetsu depends on another bunsetsu. In fact, the bunsetsus that do not have a head bunsetsu occupy 51.1% of the whole[†].

Next, we investigated inversion phenomena and dependencies over two utterance units. 256 inversions, providing 4.2% as the appearance ratio to a turn, are in this data. This fact means that the inversion phenomena can not be ignored in spoken language processing. About 85.2% of inversions appear at the last bunsetsu. On the other hand, 92 dependencies, providing 5.4% of 1,362 turns consisting of more than two units, are over two utterance units. Therefore, we can conclude that utterance units are not always sufficient as parsing units of spoken language.

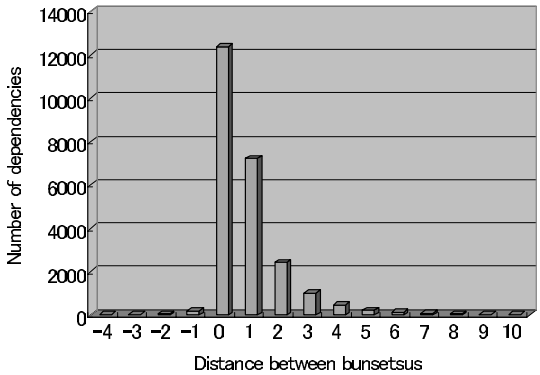

Fig. 4 Distance between dependencies and their frequencies.

Furthermore, we investigated the distance between dependencies. Figure 4 shows the relation of the distance between dependencies and its frequencies. Here, the distance between dependencies means the number of bunsetsus between the dependent bunsetsu and the head bunsetsu. Moreover, that the distance of dependency is 0 means that the dependency do not have a head bunsetsu.

3. Stochastic Dependency Parsing

Our method provides the most plausible dependency analysis for each spoken language utterance unit by relaxing syntactic constraints and utilizing stochastic information acquired from a large-scale spoken dialogue corpus. In this paper, we define a parsing unit as one turn according to the result of our investigation described in Sect. 2.2.

3.1 Dependency Structural Constraints

As Sect. 1 has already pointed out, most conventional techniques for Japanese dependency parsing have assumed three syntactic constraints. Since phenomena that hardly occur in written language appear frequently in spoken language, the actual dependency structure does not satisfy such the constraints. Our method relaxes the constraints for the purpose of robust dependency parsing. That is, our method considers that the bunsetsus, which do not have any head bunsetsu, such as fillers and hesitations, depend on themselves (relaxing the constraint that each bunsetsu depends on only one

[†]Generally, only the last bunsetsu of a sentence has no head bunsetsu in Japanese written language. Only one bunsetsu of the same kind also exits in a conversational turn, as which a parsing unit is defined in this paper. In the meaning, the number of the bunsetsus, which depend on no head bunsetsu and do not correspond to such the last bunsetsu, can be calculated as 6,306, occupying 26.0% of all bunsetsus.

other bunsetsu). Moreover, we permit that a bunsetsu depends on its left-side bunsetsu to cope with the inversion phenomena (relaxing the constraint that dependencies are directed from left to right).

3.2 Stochastic Dependency Parsing of Spoken Japanese

In our method, a sequence of bunsetsus for which a morphological analysis and bunsetsu segmentation are provided is considered as an input. For a sequence of bunsetsus, B ($= b_1 \cdots b_n$), the method identifies the dependency structure S.

The conventional methods of dependency parsing for a written language have assumed the above three syntactic constraints. Considering that there exist frequent inversions, fillers, hesitations and slips in spoken language, we established that a dependency structure fulfills only one constraint: dependencies do not cross each other[†]. However, we consider the other two constraints by reflecting the stochastic information.

Assuming that each dependency is independent, the $P(S|B)$ can be calculated as follows:

$$P(S|B) = \prod_{i=1}^{n} P(b_i \overset{rel}{\rightarrow} b_j|B), \tag{1}$$

where $P(b_i \overset{rel}{\rightarrow} b_j|B)$ is the probability that a bunsetsu b_i depends on a bunsetsu b_j when the sequence of bunsetsus B is provided. The parameter S, which maximizes the conditional probability $P(S|B)$, is regarded as the dependency structure of B and identified by dynamic programming (DP). Here, note that we use $P(b_n \overset{rel}{\rightarrow} b_j|B)$, the dependency probability for the last bunsetsu, which was not used in the usual methods for written language [5], [11], to calculate $P(S|B)$.

Next, we explain the calculation of $P(b_i \overset{rel}{\rightarrow} b_j|B)$. Our method calculates the plausibility of the dependency structure by utilizing the stochastic information. The following attributes are used for that.

- Dependent bunsetsu b_i:
  - Basic form of the rightmost independent word of b_i: h_i
  - Part-of-speech of the rightmost independent word of b_i: t_i
  - Type of dependency of b_i: r_i
  - Location of b_i: l_i
- Head bunsetsu b_j:
  - Basic form of the rightmost independent word of b_j: h_j
  - Part-of-speech of the rightmost independent word of b_j: t_j
- Distance between b_i and b_j: d_{ij}
- Number of pauses between b_i and b_j: p_{ij}

Here, if a dependent bunsetsu has one or more ancillary words, the type of dependency is the lexicon, part-of-speech and conjugated form of the rightmost ancillary word, and if not so, it is the part-of-speech and conjugated form of the rightmost morpheme. Table 2 shows several examples of the types of dependencies. Then it is permissible that d_{ij} takes a minus value to treat inversion phenomena. Moreover, the location attribute l_i indicates whether it is the last one of the turn, and is used for calculating the probability of the inversion. This is based on the observation that most inverse phenomena tend to appear at the end of the turn, as Sect. 2.2 indicates.

Table 2 Examples of the types of dependencies.

Dependent bunsetsu	Type of dependency
大きい (big)	"adjective"-"adnominal form"
買える (can buy)	"verb"-"adnominal form"
ないかな (Is there?)	"setence-final particle"-"な"
そこの (there)	"adnominal particle"-"の"
電話 が (telephone)	"case particle"-"が"
近くに (near)	"case particle"-"に"
ちょっと (briefly)	"adverb"
コンビニ (convenience store)	"noun"
えーと (well)	none
そ (hesitation expression)	none

By using the above attributes, the conditional probability $P(b_i \overset{rel}{\rightarrow} b_j|B)$ is calculated as follows:

$$\begin{aligned} P(b_i \overset{rel}{\rightarrow} b_j|B) &\cong P(b_i \overset{rel}{\rightarrow} b_j|h_i, h_j, t_i, t_j, r_i, d_{ij}, p_{ij}, l_i) \\ &= \frac{C(b_i \overset{rel}{\rightarrow} b_j, h_i, h_j, t_i, t_j, r_i, d_{ij}, p_{ij}, l_i)}{C(h_i, h_j, t_i, t_j, r_i, d_{ij}, p_{ij}, l_i)}. \end{aligned} \tag{2}$$

Note that C is a cooccurrence frequency function. The probability of a bunsetsu not having a head bunsetsu can also be calculated in formula (2) by considering that such a bunsetsu depends on itself (i.e. $i = j$).

In order to resolve sparse data problems which will be caused in estimating the $P(b_i \overset{rel}{\rightarrow} b_j|B)$ by using formula (2), we adopted the smoothing method described by Fujio and Matsumoto [5]: if $C(h_i, h_j, t_i, t_j, r_i, d_{ij}, p_{ij}, l_i)$ in formula (2) is 0, we estimate $P(b_i \overset{rel}{\rightarrow} b_j|B)$ by using formula (3).

$$\begin{aligned} P(b_i \overset{rel}{\rightarrow} b_j|B) &\cong P(b_i \overset{rel}{\rightarrow} b_j|t_i, t_j, r_i, d_{ij}, p_{ij}, l_i) \\ &= \frac{C(b_i \overset{rel}{\rightarrow} b_j, t_i, t_j, r_i, d_{ij}, p_{ij}, l_i)}{C(t_i, t_j, r_i, d_{ij}, p_{ij}, l_i)}. \end{aligned} \tag{3}$$

3.3 Parsing Example

The parsing example of a user's utterance sentence including a filler "えーと," a hesitation "そ," an inversion between "ないかな" and "近くに," and a pause, "えーと コンビニ ないかな ⟨*pause*⟩ そ そこの 近くに (Is there a convenience store near there?)" is as follows:

The sequence of bunsetsus of the sentence is "[えーと (well)], [コンビニ (convenience store)], [ないかな (Is

[†]This is because the phenomena that dependencies cross each other is very few.

Table 3 Probabilities that dependent bunsetsus depend on head bunsetsus

		Head bunsetsu					
		えーと	コンビニ	ないかな	そ	そこの	近くに
Dependent bunsetsu	えーと (well)	**1.00**	0.00	0.00	0.00	0.00	0.00
	コンビニ (convenience store)	0.00	**0.01**	0.40	0.00	0.00	0.00
	ないかな (Is there?)	0.00	0.00	**0.88**	0.00	0.00	0.00
	そ (hesitation)	0.00	0.00	0.00	**1.00**	0.00	0.00
	そこの (there)	0.00	0.02	0.00	0.00	**0.00**	0.75
	近くに (near)	0.00	0.00	0.80	0.00	0.00	**0.02**

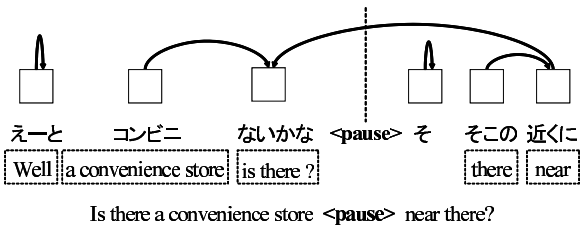

Fig. 5 Dependency structure of "[えーと (well)], [コンビニ (convenience store)], [ないかな (Is there?)], ⟨*pause*⟩, [そ], [そこの (there)], [近くに (near)]".

there?)], ⟨*pause*⟩, [そ], [そこの (there)], [近くに (near)]." The types of dependencies of bunsetsus and the dependency probabilities between bunsetsus are shown in Tables 2 and 3, respectively. Table 3 expresses that, for instance, the probability that "コンビニ (convenience store)" depends on "ないかな (Is there?)" is 0.40. Moreover, the probability of that "えーと (well)" depends on "えーと (well)" means that the probability of that "えーと (well)" does not depend on any bunsetsu. Calculating the probability of every possible structure according to Table 3, that of the dependency structure shown in Fig. 5 becomes the maximum.

4. Parsing Experiment

To evaluate the effectiveness of our method, we designed an experiment on dependency parsing. In the experiment, we used the syntactically annotated spoken language corpus [16], which we constructed by semi-automatically providing a dependency analysis for each of the driver's utterances in the CIAIR in-car speech dialogue corpus [8]–[10].

4.1 Outline of the Experiment

We used 81 dialogues, which included 6,078 turns consisting of 24,250 bunsetsus (i.e., the average length of a turn is 4.0 bunsetsus), in our syntactically annotated spoken language corpus [16]. We performed a cross-validation experiment by dividing the entire data into dialogues. That is, we repeated the experiment, in which we used one dialogue from among 81 dialogues as the test data and the others as the learning data, 81 times. Here, we define a turn as a parsing unit according to the result of the above investigation.

We also performed dependency parsing by the following two methods to use the results as baseline measures.

- The method parses a turn by deciding that each bunsetsu (except the last bunsetsu in turn) depends on the adjacent right bunsetsu and that the last bunsetsu has no head. **(Baseline1)**
- The method parses a turn by using the proposed method, but which don't allow inversions or bunsetsus with no head. **(Baseline2)**

Table 4 Experimental result about parsing accuracy.

	for dependency	for turn
Our method	87.0% (21,089/24,250)	70.1% (4,260/6,078)
Baseline1	53.8% (13,058/24,250)	30.5% (1,853/6,078)
Baseline2	57.9% (14,049/24,250)	38.3% (2,329/6,078)

4.2 Experimental Result

Table 4 shows the average accuracy of each method for a dependency or turn. In our method, among the 24,250 dependencies in the experimental data, 21,089 were correctly parsed and the accuracy was 87.0%, which was much higher than the above two baselines. We have confirmed that the parsing accuracy of our method for spontaneously spoken Japanese language is as high as that of other methods for written language [5], [6], [11], [20].

5. Discussions

In this section, we discuss the robustness for spontaneous spoken language and the sparse data problems.

5.1 Robustness for Spontaneous Spoken Language

We focus our attention on dependencies with no head bunsetsu, dependencies directed from left to right and dependencies across utterance units, and discuss the robustness of our method for spontaneous spoken Japanese based on the experimental results described in Sect. 4.

5.1.1 Dependencies with No Head Bunsetsu

Although each bunsetsu, except the last bunsetsu, has one head bunsetsu in regular written Japanese, there are some bunsetsus that have no head bunsetsu, such as fillers or hesitations, in spoken Japanese. The bunsetsus without a head bunsetsu occupy 51.1% of the whole in the used corpus. Among these bunsetsus, numbering 12,384, 4,937 are not located right before a pause. The items of these bunsetsus are shown in Fig. 6. About 70% of the whole is fillers or hesitations. Since the corpus is constructed based on the rule

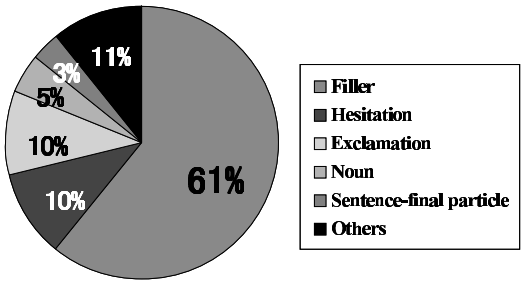

Fig. 6 Types of bunsetsus with no head bunsetsu (except the last bunsetsu of an utterance).

Table 5 Parsing result for dependencies with no head bunsetsu (except bunsetsus located right before a pause, or which is a filler or hesitation).

precision	60.4% (996/1,650)
recall	69.5% (996/1,434)

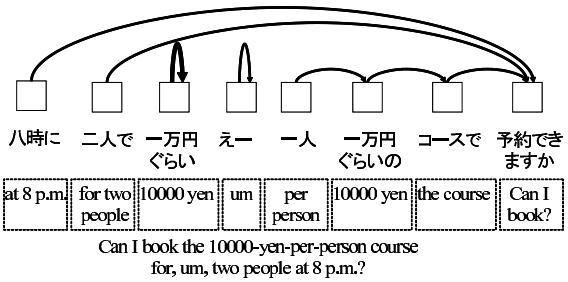

Fig. 7 Example of correctly parsed dependency with no head bunsetsu.

Table 6 Parsing result for dependencies directed from right to left.

precision	60.5% (49/ 81)
recall	19.1% (49/256)

that fillers or hesitations do not have head bunsetsus, it is not difficult to identify the head bunsetsu. Table 5 shows the experimental results for the remaining 30%. Figure 7 shows an example of the correctly parsed dependency structures with such a dependency. Among the 1,434 bunsetsus included in the 30%, 996 were correctly parsed, which means that our method can identify the dependencies with high accuracy.

5.1.2 Dependencies Directed from Right to Left

To identify inversion, our method does not assume that no dependency is directed from right to left. If the dependency parsing is performed based on the assumption that inversions exist, it becomes difficult to realize correct parsing because the search domain for identifying the head bunsetsu approximately doubles. There exist 256 dependencies directed from right to left, thus we can not always ignore them. However, since the rate is only 1% of the whole, it is not necessarily clear whether we should parse these bunsetsus.

Table 6 shows the experimental results for dependencies directed from right to left, and Fig. 8 shows an example of the correctly parsed dependency structures with such a dependency. Although the recall is not always high, the precision does exceed 60%. This means that the parsing accuracy increases by accepting that dependency is directed from right to left, that is, shows the robustness of our method for inversions. Obtaining these good results is attributed to the following two tendencies of the appearance of inversions.

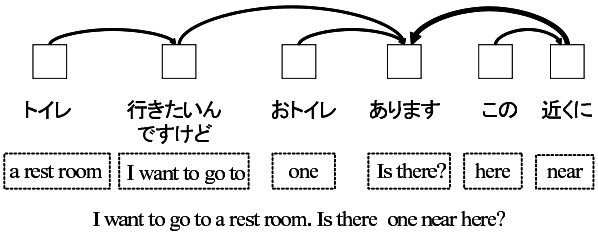

Fig. 8 Example of correctly parsed dependency directed from right to left.

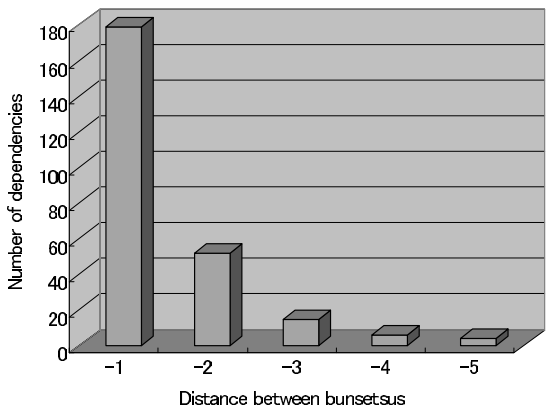

Fig. 9 Distance between bunsetsus of inversion.

One is the tendency of the location of bunsetsus. Many inversions have dependent bunsetsus that appear at the end of an utterance unit. Concretely speaking, 85.2% of inversions appear at the last bunsetsu of a turn. From the experimental results, we can see the effects of adopting the location of a dependent bunsetsu as attribute of formula (2) in consideration of the above points. Indeed, among 81 dependencies that were judged to be directed from right to left, the precision of dependencies located at the end of a turn is 75.0%.

The other is the tendency of the distance between bunsetsus. Among dependencies directed from right to left, as Fig. 9 shows, dependencies whose the distance between bunsetsus is -1 or -2 occupy 90.2% of the whole. This fact is reflected in the calculation of the dependency probability by accepting the value of d_{ij} as less than 0. In the experimental result, the precision of inversions whose distance between bunsetsus is no less than -2 is 61.0%.

5.1.3 Dependencies across Utterance Units

In spoken Japanese, it is not easy to define a grammatical unit corresponding to a sentence in written language. Although the possibility that a pause means a boundary of the unit is high, we can see a slightly different case. Thus we

Table 7 Parsing result for dependencies across utterance units.

precision	6.5% (34/521)
recall	37.0% (34/ 92)

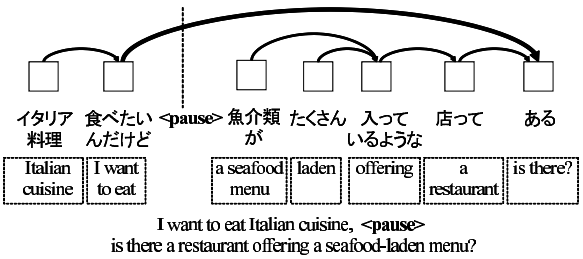

Fig. 10 Example of correctly parsed dependency across utterance units.

defined one turn as a parsing unit in the experiment. There were 92 dependencies across utterance units in the used corpus.

Table 7 shows the experimental result for dependencies across utterance units, and Fig. 10 shows an example of the correctly parsed dependency structures with such a dependency. The precision brought a remarkably low result. The result was thought to be due to the following reasons:

- The appearance frequency of those dependencies is quite low (the probability is 0.4% of the whole).
- Since the grammatical feature of dependencies across utterance units is not clear, our method does not introduce it into the probability calculation.

On the other hand, since the recall was 37.0%, we can see a certain degree of effect.

5.2 Robustness for Sparse Data Problems

As formula (2) shows, our method utilizes various attributes for robustly parsing spoken language. For the purpose of high accuracy parsing, a large-scale corpus is desired naturally. In fact, the size of the learning data in our experiment is not necessarily small as syntactically annotated data of spoken language. However, even if the data is not fully large, it can not be expected our method to be ineffective for speech understanding of spoken dialogue systems. The reasons are as follows:

1. The smoothing technique works robustly for the sparse data problems. This fact has been shown by the past studies such as Collins [3], Fujio and Matsumoto [5] though most of them have been applied to written language.
2. Most spoken dialogue systems is at least currently used for the specific task-domain such as in-car speech for which our parsing experiment was conducted. The spread of user's utterance style is not necessarily wide in the case of the limited domain.

In order to evaluate the effect for the smaller size data, we performed another experiment. We used 6,000 turns in the same 81 dialogues as Sect. 4. Among 6,000 turns, 500 turns consisting of 2,001 dependencies were used for test data. The experiments, in which the remaining 5,500 turns were used as the learning, were repeated eleven times by increasing 500 turns.

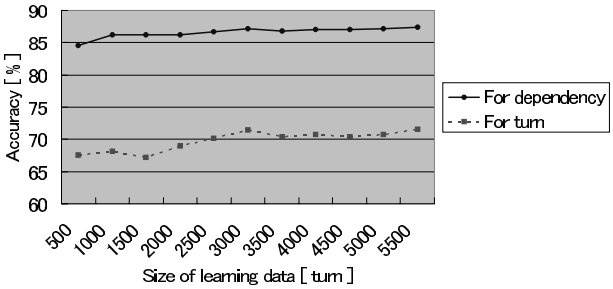

Fig. 11 Relation between accuracy and size of learning data.

Figure 11 shows the relation between the size of leaning data and parsing accuracy for a dependency or turn. This figure shows that the parsing accuracy for smaller learning data is not very lower than that for larger one.

6. Concluding Remarks

This paper proposed a method for dependency parsing of Japanese spoken language. The method can execute a robust analysis by relaxing syntactic constraints of Japanese and utilizing stochastic information. An experiment on the CIAIR in-car spoken dialogue corpus has shown our method to be effective for understanding spontaneous speech.

This experiment was conducted under the assumption that the speech recognition system has perfect performance. Since the transcript generated by a continuous speech recognition system, however, might include a lot of recognition errors, exceedingly robust parsing technologies are urgently required. To demonstrate our method as practical for automatic speech transcription, an experiment using a continuous speech recognition system will be performed in the future.

Acknowledgments

The authors would like to thank Mr. Takahisa Murase for his important contribution to their early work, and Ms. Hitomi Toyama of the Graduate School of Information Science, Nagoya University for her helpful support in correcting the syntactically annotated corpus. They wish to thank the anonymous reviewers for their helpful comments. This work is partially supported by the Grant-in-Aid for COE Research and for Young Scientists of the Ministry of Education, Science, Sports and Culture, Japan and The Tatematsu Foundation.

References

[1] J. Bear and P. Price, "Prosody, syntax, and parsing," Proc. 28th Annual Meeting of the Association for Computational Linguistics, pp.17–22, 1990.

[2] E. Charniak, "A maximum-entropy-inspired parser," Proc. 1st Conference of the North American Chapter of the Association for Computational Linguistics, pp.132–139, 2000.
[3] M. Collins, "A new statistical parser based on bigram lexical dependencies," Proc. 34th Annual Meeting of the Association for Computational Linguistics, pp.184–191, 1996.
[4] Y. Den, "A unified approach to parsing spoken natural language," Proc. 3rd Natural Language Processing Pacific Rime Symposium, pp.574–579, 1995.
[5] M. Fujio and Y. Matsumoto, "Japanese dependency structure analysis based on lexicalized statistics," Proc. 3rd Conference on Empirical Method for Natural Language Processing, pp.87–96, 1998.
[6] M. Haruno, S. Shirai, and Y. Ooyama, "Using decision trees to construct a partial parser," Proc. 17th International Conference on Computational Linguistics, pp.505–511, 1998.
[7] D. Hindle, "Deterministic parsing of syntactic nonfluencies," Proc. 21th Annual Meeting of the Association for Computational Linguistics, pp.123–128, 1983.
[8] N. Kawaguchi, S. Matsubara, K. Takeda, and F. Itakura, "Multimedia data collection of in-car speech communication," Proc. 7th European Conference on Speech Communication and Technology, pp.2027–2030, 2001.
[9] N. Kawaguchi, S. Matsubara, K. Takeda, and F. Itakura, "Multidimensional data acquisition for integrated acoustic information research," Proc. 3rd International Conference on Language Resources and Evaluation, pp.2043–2046, 2002.
[10] I. Kishida, Y. Irie, Y. Yamaguchi, M. Matsubara, N. Kawaguchi, and Y. Inagaki, "Construction of an advanced in-car spoken dialogue corpus and its characteristic analysis," Proc. 8th European Conference on Speech Communication and Technology, pp.1581–1584, 2003.
[11] T. Kudo and Y. Matsumoto, "Japanese dependency analysis based on support vector machines," Proc. 2000 Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora, pp.18–25, 2000.
[12] S. Kurohashi and M. Nagao, "Building a Japanese parsed corpus while improving the parsing system," Proc. 4th Natural Language Processing Pacific Rim Symposium, pp.451–456, 1997.
[13] S. Kurohashi and M. Nagao, "KN parser: Japanese dependency/case structure analyzer," Proc. International Workshop on Sharable Natural Language Resources, pp.48–95, 1994.
[14] K. Maekawa, H. Koiso, S. Furui, and H. Isahara, "Spontaneous speech corpus of Japanese," Proc. 2nd International Conference on Language Resources and Evaluation, no.262, pp.947–952, 2000.
[15] Y. Matsumoto, A. Kitauchi, T. Yamashita, and Y. Hirano, "Japanese morphological analysis system chasen version 2.0 manual," NAIST Techinical Report, NAIST-IS-TR99009, 1999.
[16] T. Ohno, S. Matsubara, N. Kawaguchi, and Y. Inagaki, "Spiral construction of syntactically annotated spoken language corpus, Proc. 2003 IEEE International Conference on Natural Language Proceedings and Knowledge Engineering, pp.477–483, 2003.
[17] A. Stolcke and E. Shriberg, "Statistical language modeling for speech disfluencies," Proc. International Conference on Acoustics, Speech and Signal Processing, vol.1, pp.405–408, 1996.
[18] A. Ratnaparkhi, "A liner observed time statistical parser based on maximum entropy models," Proc. 2nd Conference on Empirical Method for Natural Language Processing, pp.1–10, 1997.
[19] R.C. Rose and G. Riccardi, "Modeling disfluency and background events in ASR for a natural language understanding task," Proc. International Conference on Acoustics, Speech and Signal Processing, vol.1, pp.341–344, 1999.
[20] K. Uchimoto, S. Sekine, and K. Isahara, "Japanese dependency structure analysis based on maximum entropy models," Proc. 9th European Chapter of the Association for Computational Linguistics, pp.196–203, 1999.

Tomohiro Ohno received the B.S. degree in information engineering from Nagoya University, in 2003. Since 2003, he has been a master course student at the Graduate School of Information Science, Nagoya University. His research interests include natural language processing and spoken language processing. He is a member of the IPSJ.

Shigeki Matsubara received the B.E. degree in electrical and computer engineering from the Nagoya Institute of Technology, in 1993, and the M.E. degree and the Dr. of Engineering degree in information engineering from Nagoya University, in 1995, and 1998, respectively. He was a Research Fellow of the JSPS from 1996 to 1998, and a Research Associate from 1998 to 2002 at the Faculty of Language and Culture, Nagoya University. Since 2002, he has been an Associate Professor of the Information Technology Center, Nagoya University. His research interests include natural language processing, spoken language processing, and digital library. He is a member of the ACM, the IPSJ, the JSAI, the NLP, and the JAIS.

Nobuo Kawaguchi received his B.S. degree, M.S. degree, and a Dr. of Engineering degree from Nagoya University in 1990, 1992, and 1997, respectively. He has been an Assistant, a Lecturer, and an Associate Professor at Nagoya University. Since 2002, he has been an Associate Professor at the Information Technology Center, Nagoya University. His research interests include mobile computing, mobile agent technology, and multi-modal user interfaces. He is a member of the IEEE, the ACM, the IPSJ, the ASJ, the JSSST, and the JSAI.

Yasuyoshi Inagaki received his B.S. degree, an M.S. degree, and a Dr. of Engineering degree in electronics engineering from Nagoya University in 1962, 1964, and 1967, respectively. He was an Assistant Professor at the Department of Electrical Engineering, Nagoya University, a Professor at the Department of Electronics, Mie University, and a Professor at the Department of Electrical Engineering and Information Engineering, Nagoya University. Since 2003, he has been a Professor in the Faculty of Information Science and Technology, Aichi Prefectural University. His research interests include communication and computation, algebraic theory of software specification, verification and implementation, automata and language theory, artificial intelligence, and natural language processing. He is a member of IEEE (Senior Member), the ACM, the EATCS, IPSJ (Fellow), JSSST, JSAI, and ANLP.