Stochastic Dependency Parsing of Spontaneous Japanese Spoken Language

Shigeki Matsubara† Takahisa Murase‡ Nobuo Kawaguchi†
and **Yasuyoshi Inagaki‡**
†Information Technology Center/CIAIR, Nagoya University
‡Graduate School of Engineering, Nagoya University
Furo-cho, Chikusa-ku, Nagoya, 464-8601, Japan
matubara@itc.nagoya-u.ac.jp

Abstract

This paper describes the characteristic features of dependency structures of Japanese spoken language by investigating a spoken dialogue corpus, and proposes a stochastic approach to dependency parsing. The method can robustly cope with inversion phenomena and bunsetsus which don't have the head bunsetsu by relaxing the syntactic dependency constraints. 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 the probabilities. An experiment on dependency parsing for driver's utterances in CIAIR in-car spoken dialogue corpus has been made. The experimental result has shown our method to be effective for robust parsing of spoken language.

1 Introduction

With the recent advances of the continuous speech recognition technology, a considerable number of studies have been made on spoken dialogue systems. For the purpose of smooth interaction with the user, it is necessary for the system to understand the spontaneous speech. Since spoken language includes a lot of grammatically ill-formed linguistic phenomena such as fillers, hesitations and self-repairs, grammar-oriented approaches are not necessarily suited to spoken language processing. A technique for robust parsing is thus strongly required.

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 account of such the features. The conventional methods of dependency parsing have assumed the following three syntactic constraints (Kurohashi and Nagao, 1994):

1. No dependency is directed from right to left.
2. Dependencies don't cross each other.
3. Each *bunsetsu* [1] , except the last one, depends on only one bunsetsu.

As far as we have investigated the corpus, however, many spoken utterance do not satisfy these constraints because of inversion phenomena, bunsetsus which don't have the head bunsetsu, and so on. Therefore, our parsing method relaxes the first and third ones among the above three constraints, that is, permits the dependency direction from right to left and the bunsetsu which doesn't depend on any bunsetsu. The parsing results are expressed by partial dependency structures.

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 the probabilities. Several techniques for dependency parsing based on stochastic approaches have been proposed so far. Fujio and Matsumoto have used the probability based on the frequency of cooccurrence between two bunsetsus for dependency parsing (Fujio and Matsumoto, 1998). Uchimoto et al. have proposed a technique for learning the dependency probability model based on a maximum entropy method (Uchimoto *et al.*, 1999). However, since these

[1] 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 between two bunsetsus.

techniques are for written language, whether they are available for spoken language or not is not clear. As the 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 (Den, 1995).

On the other hand, our method provides the most plausible dependency structures for natural speeches by utilizing stochastic information. In order to evaluate the effectiveness of our method, an experiment on dependency parsing has been made. In the experiment, all driver's utterances in 81 spoken dialogues of CIAIR in-car speech dialogue corpus (Kawaguchi *et al.*, 2001) have been used. The experimental result has shown our method to be available for robust parsing of spontaneously spoken language.

2 Linguistic Analysis of Spontaneous Speech

We have investigated spontaneously spoken utterances in an in-car speech dialogue corpus which is constructed at the Center for Integrated Acoustic Information Research (CIAIR), Nagoya University (Kawaguchi *et al.*, 2001) The corpus contains speeches of dialogue between drivers and navigators (humans, a Wizard of OZ system, or a spoken dialogue system) and their transcripts.

2.1 CIAIR In-car Speech Dialogue Corpus

Data collection project of in-car speech dialogues at CIAIR has started in 1999 (Kawaguchi *et al.*, 2002). The project has developed a private car, and been collecting a total of about 140 hours of multimodal data such as speeches, images, locations and so on. These data would be available for investigating in-car speech dialogues.

The speech files are transcribed into ASCII text files by hand. The example of a transcript is shown in Figure 1. As an advance analysis, discourse tags are assigned to fillers, hesitations, and so on. Furthermore, each speech is segmented into utterance units by a pause, and the exact start time and end time are provided for them. The environmental information about sex (male/female), speaker's role (driver/navigator), dialogue task

```
0003 - 00:04:955-00:06:560 M:D:N:0:
じゃあ                                  & ジャー
マック                  [McDonald's]    & マック
教えてください<SB>      [Please tell me] & オシエテクダサイ<SB>
0004 - 00:08:101-00:09:952 F:0:N:I:
はい                    [Yes]           & ハイ
マクドナルドですね<SB> [McDonald's]     & マクドナルドデスネ<SB>
0005 - 00:10:665-00:14:111 F:0:N:0:
この先                  [Around here]   & コノサキ
二百メート先に          [200 meters away from here] & ニヒャクメートサキニ
マクドナルドが          [McDonald's]    & マクドナルドガ
あります<SB>            [There is]      & アリマス<SB>
```

Figure 1: Sample transcription of dialogue speech

(navigation/information retrieval/...), noise (noisy/clean) is provided for each utterance unit.

In order to study the features of in-car dialogue speeches, we have investigated all driver's utterance units of 195 dialogues. The number per utterance unit of fillers, hesitations and slips, are 0.34, 0.07, 0,04, respectively. The fact that the frequencies are not less than those of human-human conversations suggests the in-car speech of the corpus to be spontaneous.

2.2 Dependency Structure of Spoken Language

In order to characterize spontaneous dialogue speeches from the viewpoint of dependency, we have constructed a spoken language corpus with dependency structures. Dependency analyses have been provided by hand for all driver's utterance units in 81 spoken dialogues of the in-car speech corpus. The specifications of part-of-speeches and dependency grammars are in accordance with those of Kyoto Corpus (Kurohashi and Nagao, 1997), which is one of Japanese text corpora. 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 forms dependency structures independently.
- A bunsetsu whose head bunsetsu is omitted doesn't depend on any bunsetsu.
- The specification of part-of-speeches has been provided for the phrases peculiar to spoken language by adding lexical entries to the dictionary.
- We have defined one conversational turn as a unit of dependency parsing. The depen-

Table 1: Corpus data for dependency analysis

Dialogues	81
Utterance units	7,781
Conversational turns	6,078
Bunsetsus	24,993
Dependencies	11,789
Dependencies per unit	1.52
Dependencies per turn	1.94

dencies might be over two utterance units, but be not hardly over two conversational turns.

The outline of the corpus with dependency analyses is shown in Table 1. There exist 11,789 dependencies for 24,993 bunsetsus [2]. The average number of dependencies per turn is 1.94, and is exceedingly less than that of written language such as newspaper articles (about 10 dependencies). This does not necessarily mean that dependency parsing of spoken language is easy than that of written language. It is also required to specify the bunsetsu with no head bunsetsu because every bunsetsu does not depend on another bunsetsu. In fact, the bunsetsus which don't have the head bunsetsu occupy 52.8% of the whole.

Next, we investigated inversion phenomena and dependencies over two utterance units. 320 inversions, 3.8% of all utterance turns and about 0.04 times per turn, are in this data. This fact means that the inversion phenomena can not be ignored in spoken language processing. About 86.5% of inversions appear at the last bunsetsu. On the other hand, 73 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.

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 one turn as a parsing unit according to the result of our investigation described in Section 2.2

[2]The frequency of filler bunsetsus is 3,049.

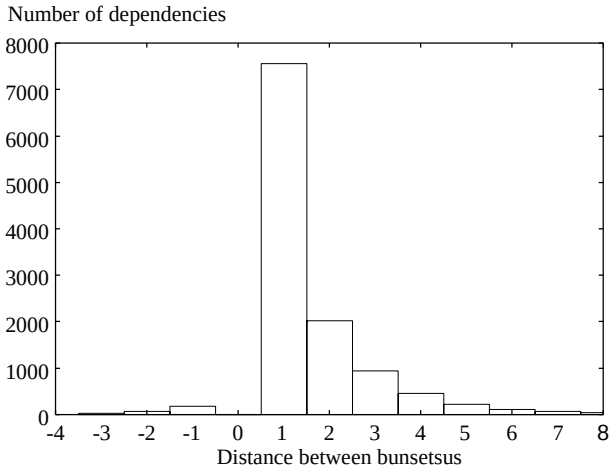

Figure 2: Distance between dependencies and its frequencies

3.1 Dependency Structural Constraints

As Section 1 has already pointed out, most conventional techniques for Japanese dependency parsing have assumed three syntactic constraints. Since the phenomena which are not hardly 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 don't have the head bunsetsu, such as fillers and hesitations, depend on themselves (relaxing the constraint that each bunsetsu depends on another only one 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].

3.2 Utilizing Stochastic Information

Our method calculates the plausibility of the dependency structure by utilizing the stochastic information. The following attributes are used for that:

- Basic forms of independent words of a dependent bunsetsu b_i and a head bunsetsu

[3]Since the phenomena that dependencies cross each other is very few, the constraint is not relaxed.

Table 2: Examples of the types of dependencies

Dependent bunsetsu	Type of dependency
denwa-ga (telephone)	case particle "ga"
mise-de (at a store)	case particle "de"
hayaku (early)	continuous form
ookii (big)	adnominal form
kaeru (can buy)	adnominal form
chotto (briefly)	adverb

b_j: h_i, h_j

- Part-of-speeches of independent words of a dependent bunsetsu b_i and a head bunsetsu b_j: t_i, t_j
- Type of the dependency of a bunsetsu b_i: r_i
- Distance between bunsetsus b_i and b_j: d_{ij}
- Number of pauses between bunsetsus b_i and b_j : p_{ij}
- Location of a dependent bunsetsu b_i: l_i

Here, if a dependent bunsetsu b_i has an ancillary word, the type of the dependencies of a bunsetsu b_i, r_i, is the lexicon, part-of-speech and conjugated form of the word, and if not so, r_i is the part-of-speech and the conjugated form of the last morpheme. Table 2 shows several examples of the types of dependencies. The location of the dependent bunsetsu means whether it is the last one of the turn or not. As Section 2 indicates, the method uses the location attribute for calculating the probability of the inversion, because most inverse phenomena tend to appear at the last of the turn.

The probability of the dependency between bunsetsus are calculated using these attribute values as follows:

$$P(i \overset{rel}{\rightarrow} j|B) = \frac{C(i \rightarrow j, h_i, h_j, t_i, t_j, r_i)}{C(h_i, h_j, t_i, t_j, r_i)} \times \frac{C(i \rightarrow j, r_i, d_{ij}, p_{ij}, l_i)}{C(r_i, d_{ij}, p_{ij}, l_i)} \quad (1)$$

Here, C is a cooccurrence frequency function and B is a sequence of bunsetsus $(b_1 b_2 \cdots b_n)$.

In the formula (1), the first term of the right hand side expresses the probability of cooccurrence between the independent words, and the second term does that of the distance between bunsetsus. The problem of data sparseness is reduced by considering these phenomena to be independent each other and separating the probabilities into two terms. The probability that a bunsetsu which doesn't have the head bunsetsu can also be calculated in formula (1) by considering such the bunsetsu to depend on itself (i.e., $i = j$). The probability that a dependency structure for a sequence of bunsetsus B is S can be calculated from the dependency probabilities between bunsetsus as follows.

$$P(S|B) = \prod_{i=1}^{n} P(i \overset{rel}{\rightarrow} j|B) \quad (2)$$

For a sequence of bunsetsus, B, the method identifies the dependency structure with "$argmax_S P(S|B)$" satisfying the following conditions:

- Dependencies do not cross each other.
- Each bunsetsu doesn't no more than one head bunsetsu.

That is, our method considers the dependency structure whose probability is maximum to be the most plausible one.

3.3 Parsing Example

The parsing example of a user's utterance sentence including a filler "eto", a hesitation "so", a inversion between "nai-ka-na" and "chikaku-ni", and a pause, "Eto konbini nai-ka-na $\langle pause \rangle$ so sono chikaku-ni (Is there a convenience store near there?)" is as follows:

The sequence of bunsetsus of the sentence is "[eto (well)],[konbini (convenience store)],[nai-ka-na (Is there?)],$\langle pause \rangle$, [so], [sono (there)], [chikaku-ni (near)]". The types of dependent of bunsetsus and the dependency probabilities between bunsetsus are shown in Table 2 and 3, respectively. Table 3 expresses that, for instance, the probability that "konbini" depends on "nai-ka-na" is 0.40. Moreover, the probability of that "eto" depends on "eto" means that the probability of that "eto" does not depend on any bunsetsu. Calculating the probability of every possible structure according to Table 3, that of the dependency structure shown in Figure 3 becomes the maximum.

Table 3: Dependency probabilities between bunsetsus

	eto	konbini	nai-ka-na	so	soko-no	chikaku-ni
eto (well)	**1.00**	0.00	0.00	0.00	0.00	0.00
konbini (convenience store)	0.00	**0.01**	0.40	0.00	0.00	0.00
nai-ka-na (Is there?)	0.00	0.00	**0.88**	0.00	0.00	0.00
so (hesitation)	0.00	0.00	0.00	**1.00**	0.00	0.00
soko-no (there)	0.00	0.02	0.00	0.00	**0.00**	0.75
chikaku-ni (near)	0.00	0.00	0.80	0.00	0.00	**0.02**

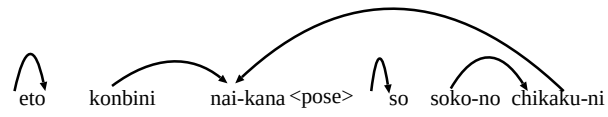

Figure 3: Dependency structure of "eto konbini nai-kana ⟨*pose*⟩ so soko-no chikaku-ni"

Table 4: Experimental result of dependency parsing

Item	(a)	(b)	(a)+(b)
Precision	82.0%	88.5%	85.5%
Recall	64.3%	83.3%	73.8%

(a): The result for 241 bunsetsus with a head
(b): The result for 240 bunsetsus with no head
(a)+(b): The result for 481 bunsetsus

4 Parsing Experiment

In order to evaluate the effectiveness of our method, an experiment on dependency parsing has been made using a corpus of CIAIR (Kawaguchi *et al.*, 2001).

4.1 Outline of Experiment

We used the same data as that for our investigations in Section 2.2. That is, among all driver's utterance units of 81 dialogues, 100 turns were used for the test data, and 5978 turns for the learning data. The test data, the average bunsetsus per turn is 4.81, consists of 481 dependencies.

4.2 Experimental Result

The results of the parsing experiment are shown partially in Figure 4. Table 4 shows the evaluation. For the parsing accuracy, both precision and recall are measured. 355 of 415 dependencies extracted by our method are correct dependencies, providing 85.5% for precision rate and 73.8% for recall rate. We have confirmed that the parsing accuracy of our method for spoken language is as high as that of another methods for written language (Fujio and Matsumoto, 1998; Uchimoto *et al.*, 1999).

Our method correctly specified 200 of 240 bunsetsus which don't have the head bunsetsu. Most of them are fillers, hesitations and so on. It became clear that it is effective to utilize the dependency probabilities for identifying them.

5 Concluding Remarks

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

This experiment has been made on the assumption that the speech recognition system has a 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 strongly required. In order to demonstrate our method to be practical for automatic speech transcription, an experiment using a continuous speech recognition system will be done.

Acknowledgement: The authors would like to thank all members of SLP Group in our laboratory for their contribution to the construction of the Japanese spoken language corpus with the dependency analysis. This work is partially supported by the Grand-in-Aid for COE

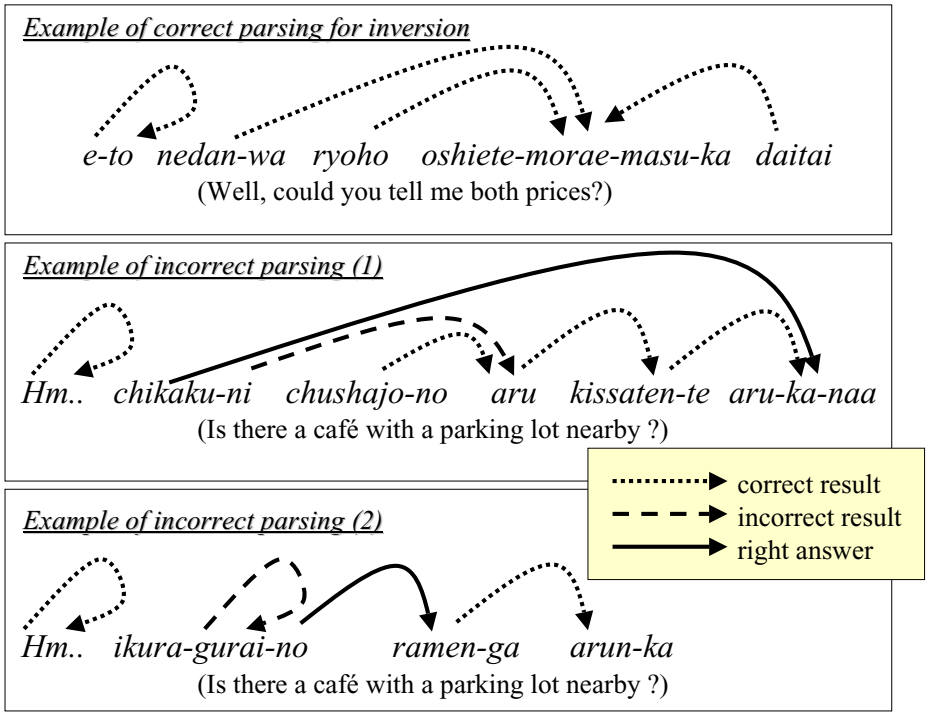

Figure 4: The results of parsing experiment (a part)

Research of the Ministry of Education, Science, Sports and Culture, Japan and Aritificial Intelligence Research Promotion Foundation.

References

Den, Y.: A Unified Approach to Parsing Spoken Natural Language, *Proceedings of 3rd Natural Language Processing Pacific Rime Symposium (NLPRS'95)*, pp. 574–579 (1995).

Fujio, M. and Matsumoto, Y.: Japanese Dependency Structure Analysis based on Lexicalized Statistics, *Proceedings of 3rd Conference on Empirical Method for Natural Language Processing (EMNLP'98)*, pp. 87–96 (1998).

Kawaguchi, N., Matsubara, S., Takeda, K., and Itakura, F.: Multi-Dimensional Data Acquisition for Integrated Acoustic Information Research, *Proceedings of 3rd International Conference on Language Resources and Evaluation (LREC2002)*, pp. 2043–2046 (2002).

Kawaguchi, N., Matsubara, S., Takeda, K. and Itakura, F.: Multimedia Data Collection of In-car Speech Communication, *Proceedings of 7th European Conference on Speech Communication and Technology (Eurospeech2001)*, pp. 2027–2030 (2001).

Kurohashi, S. and Nagao, M.: Kyoto University Text Courpus Project, *Proceedings of 3rd Conference of Association for Natural Language Processing*, pages:115–118 (1997). (In Japanese)

Kurohashi, S. and Nagao, M.: "KN Parser: Japanese Dependency/Case Structure Analyzer" *Proceedings of Workshop on Sharable Natural Language Resources*, pages:48–95 (1994).

Matsumoto, Y., Kitauchi, A., Yamashita, T. and Hirano, Y.: Japanese Morphological Analysis System Chasen version 2.0 Manual, *NAIST Techinical Report*, NAIST-IS-TR99009 (1999).

Uchimoto, K., Sekine, S. and Isahara, K.: Japanese Dependency Structure Analysis based on Maximum Entropy Models, *Proceedings of 9th European Chapter of the Association for Computational Linguistics (EACL'99)*, pp. 196–203 (1999).