Robust Dependency Parsing of Spontaneous Japanese Speech and Its Evaluation

Tomohiro Ohno†, *Shigeki Matsubara*‡, *Nobuo Kawaguchi*‡ *and Yasuyoshi Inagaki*§

† Graduate School of Information Science, Nagoya University
‡ Information Technology Center/CIAIR, Nagoya University
§ Faculty of Information Science and Technology, Aichi Prefectural University
Furo-cho, Chikusa-ku, Nagoya, 464-8603 Japan
ohno@el.itc.nagoya-u.ac.jp

Abstract

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 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. As a result of an experiment, the parsing accuracy provided 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.

1. Introduction

In order to develop a user-friendly spoken dialogue system, a technique for robustly understanding spoken language is strongly required. Although most conventional spoken dialogue systems are only treating words or expressions decided beforehand, if the system tries to execute a richer spoken dialogue, analyzing the language structure of spontaneous speech is essential. On the other hand, spontaneously spoken Japanese includes a lot of grammatically ill-formed linguistic phenomena such as fillers, hesitations, inversions, etc., which the conventional dependency parsing method for written language can not parse correctly.

This paper proposes a method of robust dependency parsing using stochastic information. The usual methods of Japanese dependency parsing have assumed the following three syntactic constraints[4]:

1. No dependency is directed from right to left.

2. Dependencies do not 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 utterances do not satisfy these constraints because of inversion phenomena, bunsetsus which do not have the head bunsetsu, and so on[5]. Therefore, our parsing method relaxes the above three constraints, that is, permits the dependency direction from right to left and the bunsetsu which does not depend on any bunsetsu.

In this paper, we also evaluate the availability and robustness of our method. Especially, we report how effectively the method can parse bunsetsus which 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 stochastic dependency parsing of spoken language. Section 3 presents the parsing experiment. The discussion about our experiment is reported in Section 4.

2. Robust 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 don't cross each other. However, we consider the other two constraints by reflecting the stochastic information.

[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.

Table 1: Experimental result about parsing accuracy

	for dependency	for turn
average accuracy	87.0% (21,089/24,250)	70.1% (4,260/6,078)

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

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

where $P(b_i \stackrel{rel}{\rightarrow} b_j|B)$ 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).

Next, we explain the calculation of $P(b_i \stackrel{rel}{\rightarrow} b_j|B)$. First, the basic form of independent words in a dependent bunsetsu is represented by h_i, its part-of-speech t_i, type of dependency r_i, and the basic form of the independent word in a head bunsetsu h_j, its part-of-speech t_j. Furthermore, the distance between bunsetsus is denoted as d_{ij}, the number of pauses between them p_{ij}, and the location of the dependent bunsetsu l_i. Here, if a dependent bunsetsu has an ancillary word, the type of the dependency is the lexicon, part-of-speech and conjugated form of that ancillary word, and if not so, it is the part-of-speech and conjugated form of the last morpheme. Then it is allowed that d_{ij} takes a minus value to treat inversion phenomena. Moreover, the dependent bunsetsu's location indicates whether it is the last one of the turn. 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[5].

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

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

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. $i = j$).

3. Parsing experiment

In order to evaluate the effectiveness of our method, we made an experiment on dependency parsing. In the experiment, we used the syntactically annotated spoken language corpus[6] which we constructed by semi-automatically providing dependency analysis for each of the driver's utterances in CIAIR in-car speech dialogue corpus[2, 3].

3.1. Outline of the experiment

We used 81 dialogues, which included 6,078 turns consisting of 24,250 bunsetsus (i.e., the average length of turn is 4.0 bunsetsus), in our syntactically annotated spoken language corpus[6]. We performed the cross validation experiment by dividing the entire data into dialogues. That is, we repeated the experiment, in which we used one dialogue 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 our investigation[5].

3.2. Experimental result

Table 1 shows the average accuracy for dependency or turn. Among the 24,250 dependencies in the experimental data, 21,089 were correctly parsed and its accuracy was 87.0%. We have confirmed that the parsing accuracy of our method for spontaneously spoken Japanese language is as high as that of another methods for written language[1, 7].

4. Discussions

In this section, we focus attention on dependencies with no head bunsetsu, dependencies directed from left to right and dependencies over utterance units, and discuss the robustness of our method for spontaneous spoken Japanese based on the experimental results described in section 3.

4.1. Dependencies with no head bunsetsu

Although each bunsetsu, except the last bunsetsu, has one head bunsetsu in usual written Japanese, there are some bunsetsus, which has no head bunsetsu, such as fillers or hesitations, in spoken Japanese. Figure 1 shows an example of the dependency structure with such the dependency. The bunsetsus which do not have the head bunsetsu occupy 51.1% of the whole in the used corpus. Among these bunsetsus, whose number is 12,384, 4,937 bunsetsus are not located right before pause. The items of these bunsetsus are shown in Fig. 2. About 70% of the whole is fillers or hesitations. Since the corpus are constructed based on the rule that fillers or hesitations do

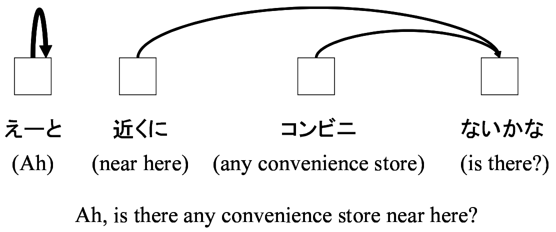

Figure 1: Example of dependency with no head bunsetsu

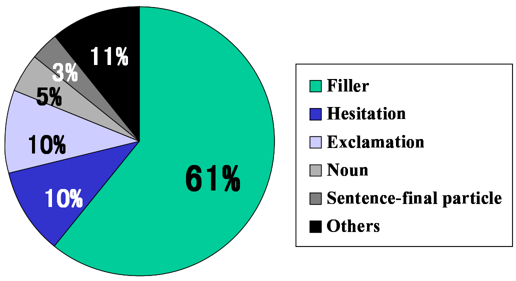

Figure 2: Types of bunsetsus with no head bunsetsu (except the last bunsetsu of an utterance)

Table 2: Parsing result for dependencies with no head bunsetsu (except bunsetsus which is located right before pause, or which is fillers or hesitation)

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

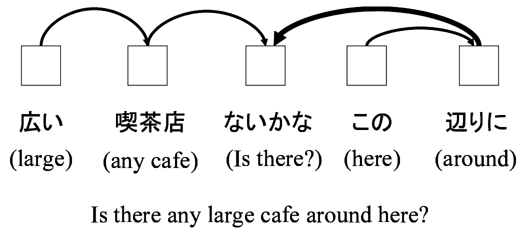

Figure 3: Example of dependency directed from right to left

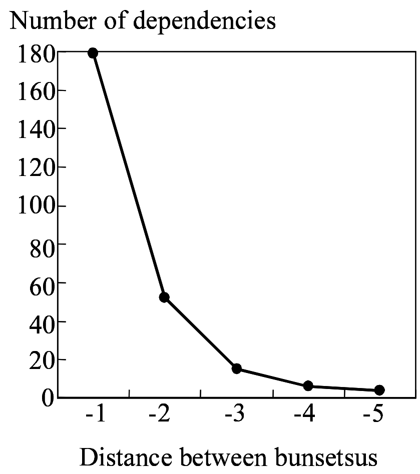

Figure 4: Distance between bunsetsus of inversion

not have the head bunsetsus, it is not difficult to identify the head bunsetsu. Consequently, Table 2 shows the experimental results for the remaining 30%. Among 1,434 bunsetsus which were included in the 30%, 996 were correctly parsed. Thus it means that our method can identify the dependencies with high accuracy.

4.2. Dependencies directed from right to left

In order to identify inversion, our method does not assume that no dependency is directed from right to left. An example of inversion is shown in Fig. 3. 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 spreads about twice. There existed 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 or not.

Table 3 shows the experimental results for dependencies directed from right to left. Although the recall is not always high, the precision is over 60%. This means that the parsing accuracy increase by accepting that dependency is directed from right to left, furthermore shows the robustness of our method for inversions. Obtaining these good results is attributed to the following two tendencies about the appearance of inversions.

One is the tendency about the location of bunsetsus. Many of inversions have the dependent bunsetsus which appears at the last 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 that we adopt the location of a dependent bunsetsu as the attribute of the formula (2) in consideration of the above things. Indeed, among 81 dependencies which were judged to be directed from right to left, the precision of dependencies which is located at the last of a turn is 75.0%.

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

Table 3: Parsing result for dependencies directed from right to left

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

4.3. Dependencies over 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 little different case. Thus we defined one turn as a parsing unit in the experiment. There were 92 dependencies over utterance units in the used corpus. Figure 5 shows an example of such the dependency.

Table 4 shows the experimental result for dependencies over utterance units. The precision brought a remarkable low result. The result was thought to be caused by the following reason:

- The appearance frequency of those dependencies is not many (the probability is 0.4% of the whole).
- Since the grammatical feature of dependencies over 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 the certain degree of effect.

5. Conclusions

In this paper, we have proposed a method of robust dependency parsing of Japanese speech using a large-scale spoken language corpus, and also evaluated the availability and robustness of the method using spontaneously spoken dialogue sentences. In the result, we have confirmed that our method can efficiently parse dependencies with no head bunsetsu, dependencies directed from right to left and dependencies over utterance units.

6. Acknowledgements

The authors would like to thank Ms. Hitomi Toyama of Nagoya University for her helpful support in correcting the dependency corpus. This work is partially supported by the Grant-in-Aids for COE Research and for Young Scientists of the Ministry of Education, Science, Sports and Culture, Japan and The Tatematsu Foundation.

7. References

[1] 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.

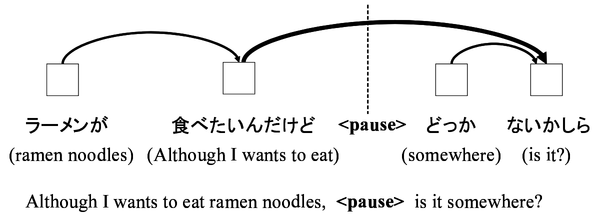

Figure 5: Example of dependency over utterance units

Table 4: Parsing result for dependencies over utterance units

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

[2] 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*(Eurospeech'01), pp. 2027-2030, 2001.

[3] Kishida, I., Irie, Y., Yamaguchi, Y., Matsubara, M., Kawaguchi, N. and Inagaki, Y., "Construction of an Advanced In-Car Spoken Dialogue Corpus and its Characteristic Analysis," *Proceedings of 8th European Conference on Speech Communication and Technology*(Eurospeech'03), pp. 1581-1584, 2003.

[4] Kurohashi, S. and Nagao, M., "Building a Japanese Parsed Corpus while Improving the Parsing System," *Proceedings of 4th Natural Language Processing Pacific Rim Symposium*(NLPRS'97), pp. 451-456, 1997.

[5] Matsubara, S., Murase, T., Kawaguchi, N. and Inagaki, Y., "Stochastic Dependency Parsing of Spontaneous Japanese Spoken Language," *Proceedings of 19th International Conference on Computational Linguistics*(COLING'02), Vol.1, pp. 640-645, 2002.

[6] Ohno, T., Matsubara, S., Kawaguchi, N. and Inagaki, Y., "Spiral Construction of Syntactically Annotated Spoken Language Corpus," *Proceedings of 2003 IEEE International Conference on Natural Language Processing and Knowledge Engineering*(NLP-KE'03), pp. 477-483, 2003.

[7] 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.