Hierarchical Coordinate Structure Analysis for Japanese Statutory Sentences Using Neural Language Models

Takahiro Yamakoshi†, Tomohiro Ohno††, Yasuhiro Ogawa†††, Makoto Nakamura†††† and Katsuhiko Toyama††††

We propose a method for analyzing the hierarchical coordinate structure of Japanese statutory sentences using neural language models (NLMs). Our method deterministically identifies hierarchical coordinate structures according to their rigorously defined descriptive rules. In addition, our method identifies all conjuncts in each coordinate structure using NLM-based scoring. Furthermore, it does not rely on any training data labeled with coordinate structures. An experiment demonstrates that our method drastically outperforms an existing method for Japanese statutory sentences.

Key Words: Coordination Analysis, Hierarchical Coordinate Structure, Neural Language Model, Japanese Statutory Sentence

1 Introduction

Japanese statutory sentences are not easy for non-experts to read. One reason for this difficulty is the frequent appearance of hierarchical coordinate structures. For example, the sentence in Figure 1 contains four coordinate structures that compose a three-layered hierarchical coordinate structure. Such complex hierarchical coordinate structures also degrade the quality of automatic statutory document processing. Therefore, the development of a high-performance technique for coordination analysis is desired in statutory document processing, such as reading assistance with statutory sentences (Yamada and Shimazu 2006) and the automatic construction of a legal term thesaurus (Hagiwara, Ogawa, and Toyama 2009).

Various methods have been proposed for coordination analysis (e.g., (Agarwal and Boggess 1992; Kurohashi and Nagao 1994; Resnik 1999; Goldberg 1999; Chantree, Kilgarriff, de Roeck, and Willis 2005; Dubey, Keller, and Sturt 2006; Kawahara and Kurohashi 2008; Roh, Lee, Choi, Kwon, and Kim 2008; Hara, Shimbo, Okuma, and Matsumoto 2009; Hanamoto, Matsuzaki, and Tsujii 2012; Ficler and Goldberg 2016)). However, we cannot expect these methods to work well

---

† Graduate School of Informatics, Nagoya University
†† Graduate School of Advanced Science and Technology, Tokyo Denki University
††† Information Technology Center, Nagoya University
†††† Department of Engineering, Niigata Institute of Technology
for Japanese statutory sentences because they fail to adequately consider the characteristics of such sentences in Japanese. In fact, even a famous Japanese parser, KNP (v3.01) (Kawahara and Kurohashi 2006), could identify coordinate structures in Japanese statutory sentences at only 26 points of F-measure (Matsuyama, Shirai, and Shimazu 2012).

At present, only one method is specialized for Japanese statutory sentences (Matsuyama et al. 2012) (hereafter, the conventional method). This method deterministically identifies the scope of all conjuncts in coordinate structures based on the descriptive rules unique to Japanese statutory sentences (Ishige 2004; Oshima 2005; Maeda 1983). However, the analysis performance of this method is less than sufficient. One reason is that it has difficulty identifying conjuncts whose word length is very different from that of their adjacent conjuncts. This is because this method uses scoring based on one-to-one word alignment as a clue to identifying conjuncts.

In this paper, we propose a new method for identifying hierarchical coordinate structures in Japanese statutory sentences using neural language models (NLMs) (Bengio, Ducharme, Vincent, and Jauvin 2003). Our method inherits the deterministic analysis strategy specialized for Japanese statutory sentences from the conventional method. In addition, it uses NLMs instead of one-to-one word alignment. Since our method transforms each conjunct candidate into a single fixed-length vector using NLMs, it can identify conjuncts without being affected by their length. Here, as with the conventional method, our method does not rely on any training data labeled with coordinate structures, since our NLMs are trained by raw text. In an experiment using Japanese statutory sentences, we confirmed our method substantially outperforms the conventional one.

This paper is structured as follows. In the next section, we introduce related work that focuses on coordinate analysis. We describe coordinate structures that are unique to Japanese statutory sentences in Section 3. We outline the conventional method in Section 4. After describing the basic concept of the neural language model in Section 5, we explain our method mainly
by comparing it with the conventional method in Section 6. The evaluation experiment and a discussion are given in Sections 7 and 8, respectively. Finally, we summarize and conclude this paper in Section 9.

2 Related Work

As described in Section 1, among the many existing methods for coordination analysis, only the method proposed by Matsuyama et al. (2012) can deal with the coordinate structures unique to Japanese statutory sentences. Therefore, we propose a new method based on that previous method, refining its weak points. Some other works have shown the effect of considering domain-specific knowledge. For example, Roh et al. (2008) proposed a method specialized for English patent documents using regular expressions.

As is well known, a number of methods have been developed for domain-independent coordination analysis. Among them, many methods, especially the older ones, focus on identifying only two conjuncts adjacent to a coordinator (Agarwal and Boggess 1992; Resnik 1999; Goldberg 1999; Chantree et al. 2005; Dubey et al. 2006; Ficler and Goldberg 2016). Since coordinate structures in Japanese statutory sentences often have three or more conjuncts, we need to identify all of them.

Of course, there are also many methods that can cope with three or more conjuncts (Kurohashi and Nagao 1994; Kawahara and Kurohashi 2008; Hara et al. 2009; Hanamoto et al. 2012). However, most of these methods use previously or simultaneously obtained syntactic information (Kawahara and Kurohashi 2008; Hanamoto et al. 2012), or are based on one-to-one word alignment similar to the conventional method (Kurohashi and Nagao 1994; Hara et al. 2009). Since a statutory sentence is often very long and syntactically complex, its parsing is not easy. Therefore, we propose a method that does not require syntactic information.

Finally, as is the case with our research, Ficler and Goldberg (2016) also proposed a method based on neural networks; however, it takes a different approach from our NLMs. Specifically, this method has three main differences from ours. First, it cannot handle coordinate structures with three or more conjuncts. Second, it requires syntactic trees. Third, it requires a great amount of training data labeled with coordinate structure information to tune the parameters of the neural networks. Since few resources contain such information in the Japanese statutory domain, we propose an unsupervised method using NLMs.

---

1 In our experimental data, 17.6% of coordinate structures have three or more conjuncts.
3 Coordinate Structures Unique to Japanese Statutory Sentences

In this section, we explain the coordinate structures unique to Japanese statutory sentences according to Japanese legislation-drafting manuals, which were established more than 100 years ago, e.g., (Ishige 2004; Oshima 2005; Maeda 1983). These manuals define usage of the two pairs of Japanese coordinators “matawa (or_H)” and “moshikuwa (or_L),” and “oyobi (and_L)” and “narabini (and_H),” which constitute hierarchical coordinate structures to disambiguate the interpretations. Furthermore, they define usage of two coordinators “sonota (other_1)” and “sonotano (other_2),” which have similar but slightly different meanings.

3.1 “matawa (or_H)” and “moshikuwa (or_L)”

Both “matawa” and “moshikuwa” are generally translated as “or” in English. In general Japanese sentences, these two coordinators are used interchangeably. However, in Japanese statutory sentences, “matawa” is strictly used for the highest coordinate structure, while “moshikuwa” is strictly used for other coordinate structure. Therefore, in this paper, we translate “matawa” as “or_H” and “moshikuwa” as “or_L” in English to distinguish them. Figure 1 demonstrates the usage of the two coordinators in a specific Japanese statutory sentence. Note that “matawa” is used for “or”-wise coordinate structures without hierarchy.

3.2 “oyobi (and_L)” and “narabini (and_H)”

Both “oyobi” and “narabini” are generally translated as “and” in English, and are also used interchangeably in general Japanese sentences. However, in Japanese statutory sentences, “oyobi” is strictly used for the lowest coordinate structure, while “narabini” is strictly used for other coordinate structure. In this paper, we translate “oyobi” as “and_L” and “narabini” as “and_H” in English. Figure 2 illustrates the usage of the two coordinators. Note that “oyobi” is used for “and”-wise coordinate structures without hierarchy.

![Fig. 2 Japanese statutory sentence with “oyobi (and_L)” and “narabini (and_H)”](image-url)
3.3 “sonota (other1)” and “sonotano (other2)”

Both “sonota” and “sonotano” are generally translated as “other,” “any other,” “such as,” and so on in English, and they are also used interchangeably in general Japanese sentences. However, in Japanese statutory sentences, these coordinators have different meanings, and thus they represent exact legal effects of the statutory sentences. In this paper, we translate “sonota” as “other1” and “sonotano” as “other2.”

In view of the differences in meaning, the two coordinators are considered to constitute different types of coordinate structures (Ogawa, Yamada, Kato, and Toyama 2011). Concretely, the phrase followed by “sonota (other1)” is a conjunct along with the preceding phrases. In Figure 3, the phrase “hito no chikaku ni yottewa ninshikisuru koto ga dekinai hoshiki (device not recognizable to the human senses)” after the coordinator “sonota (other1)” is a conjunct along with “denshiteki hoshiki (electronic device)” and “jikiteki hoshiki (magnetic device).”

On the other hand, the phrase succeeding “sonotano (other2)” is not a conjunct but just a hypernym of the preceding conjuncts. In addition, “sonota” in “sonotano” is also a conjunct. Here, “sonotano (other2)” is a compound word that consists of two morphemes “sonota (other)” and “no (of).” In Figure 4, the phrase “kinmujoken (working conditions)” after the coordinator “sonotano (other2)” is the hypernym of the preceding three conjuncts “kyuyo (salaries),” “kinmujikan (working hours),” and “sonota.”
4 Conventional Method for Japanese Statutory Sentences

In this section, we explain the conventional method for Japanese statutory sentences (Matsuyama et al. 2012), describing the causes of the performance decrement that has been refined in our proposed method.

Figure 5 shows the processing flow of the conventional method. The method identifies all of the coordinate structures $crd^i$ $(1 \leq i \leq N)$ in a sentence, where $N$ is the number of coordinators included in the sentence.

The method assumes that any coordinate structure in Japanese statutory sentences can be formalized with the following extended regular expression:

$$crd^i = c^i_{n_i} \cdot " \cdot c^i_{n_i-1} \cdot " \cdot \ldots \cdot c^i_1 \cdot " \cdot k^i \cdot " \cdot k^i \cdot " \cdot c^i_0; \quad (1)$$

where $c^i_j$ $(0 \leq j \leq n_i, \ n_i \geq 1)$ is the $j$-th conjunct that constitutes $crd^i$, $k^i$ represents the coordinator of $crd^i$, and “,” is a comma in Japanese. “.” represents the string concatenation operator, and “X?” means that “X” occurs at most once. Note that $k^i$ occurs only between $c^i_1$ and $c^i_0$. Figure 1 also shows that each coordinate structure is expressed in accordance with Eq. (1).

The method identifies coordinate structures deterministically, that is, it identifies each con-
junct and coordinate structure in a predetermined order and does not modify identified conjuncts or coordinate structures.

In the following subsections, we describe each process of the method in detail.

4.1 Coordinator Extraction and Ranking

All targeted coordinators shown in Table 1 are extracted from an input sentence. Two types of new coordinators “*to* (and)” and “*ya* (and),” which are also found in general Japanese sentences, have no specific usage in Japanese Statutory Sentences. Another new coordinator, “*katsu* (and),” is used for connecting semantically inseparable conjuncts in Japanese statutory sentences (Ishige 2004; Oshima 2005; Maeda 1983).

The analysis order of each of the extracted coordinators is decided based on two rules: (1) a coordinator with a smaller priority number in Table 1 comes first (Priority Rule), and (2) a coordinator that appears earlier in a sentence comes first among coordinators with the same priority (Position Rule).

4.2 Candidate Extraction

Candidates of a conjunct $c_{ij}$ are extracted as a set $C_{ij}$. Although the way to extract them varies with the value of $j$, there are two common conditions independent of $j$. First, no candidate includes a comma or any unprocessed coordinator. Second, if a candidate includes any part of an already identified structure, the candidate has to include the whole of that structure.

In the case of $C_{i1}$, the leftmost word of every candidate must be the leftmost word of a *bunsetsu*, and the rightmost word must be a word (except a comma) just before the coordinator $k^i$.

In the case of $C_{i0}$, the leftmost word of every candidate must be a word (except a comma) just after $k^i$. The rightmost word of every candidate must have the same POS as any one of the rightmost words of candidates of $c_{i1}$ and appear last in each *bunsetsu*. If the rightmost word

| Priority | Coordinator |
|----------|-------------|
| 1        | *oyobi* (及び; and$_L$) *moshikuwa* (若しくは; or$_L$) |
| 2        | *narabini* (並びに; and$_H$) *matawa* (又は; or$_H$) |
| 3        | *sonota* (その他; other$_1$) *katsu* (かつ; and) *to* (と; and) *ya* (や; and) |

*Table 1* Coordinators targeted in conventional method and their priority

---

2 *Bunsetsu* is a linguistic unit in Japanese that roughly corresponds to a basic phrase in English. A *bunsetsu* consists of one independent word and zero or more ancillary words.
is a noun, **Semantic Similarity** between the candidates of \( c_{i0} \) and \( c_{i1} \) is calculated based on a Japanese thesaurus (Shirai, Ooyama, Ikehara, Miyazaki, and Yokoo 1998), and the highest three candidates are chosen.

In the case of \( C_j^i \) (\( j \geq 2 \)), the leftmost word of every candidate must be a leftmost word of each *bunsetsu*, and the rightmost word must be a word (except a comma) before the leftmost word of \( c_{j-1}^i \).

For example, in Figure 1, the two candidate sets \( C_1^1 \) and \( C_0^1 \) for the coordinator \( k^1 \) will be \{“kokkai no ryoin (both Houses of the Diet)”, “ryoin (both Houses)”\} and \{“ichiin (either House)”\}, respectively.

### 4.3 Conjunct Identification

The two conjuncts \( c_{i1} \) and \( c_{i0} \) are identified simultaneously. Concretely, the method chooses a pair from \( C_i^1 \times C_i^0 \), which has the highest similarity between the two conjunct candidates. As for \( c_j^i \) (\( j \geq 2 \)), the similarity between each candidate in \( C_j^i \) and already identified \( c_{j-1}^i \) is calculated, and then the highest one is selected from \( C_j^i \).

The similarity between the two conjunct candidates (hereafter, **Conjunct Similarity**) is calculated using one-to-one word alignment between the candidates. Concretely, the Conjunct Similarity is based on the following two criteria: (1) ratio of words that correspond to a word in the paired candidate and (2) the sum of similarity between two words that correspond to each other. Here, similarity between two words is calculated based on their POSs and Semantic Similarity. The most appropriate word alignment is calculated by using dynamic programming.

However, this calculation method, which uses one-to-one word alignment, is weak in identifying a conjunct whose adjacent conjunct has a different word length. Figure 6 shows an example of coordinate structure that the conventional method tends to analyze incorrectly. The correct conjuncts of this coordinate structure are “*shinsei wo shi yo to suru mono* (persons planning to file applications)” and “*shinsei sha* (applicants),” but their word alignment does not work well, as shown in Figure 7(a). The words “*wo*”, “*shi*”, “*yo*”, “*to*”, “*suru*” (planning to) have

![Fig. 6 Coordinate structure with unequal length of conjuncts](image-url)
no correspondence because no word can have correspondences to multiple words. Since the conventional method imposes a penalty on words that do not have correspondence, the Conjunct Similarity between unequal-length conjuncts tends to become low. On the other hand, as shown in Figure 7(b), there is just one lone word in the word alignment between the two conjuncts “mono (persons)” and “shinsei sha (applicants),” and thus their Conjunct Similarity becomes larger than that of the two conjuncts “shinsei wo shi yo to suru mono (persons planning to file applications)” and “shinsei sha (applicants).” As a result, the conventional method incorrectly chooses the two conjuncts “mono (persons)” and “shinsei sha (applicants).”

4.4 Further Conjunct Existence Judgment

After identifying conjuncts $c_j^i$, $c_{j-1}^i$, ..., $c_0^i$, the method judges that a next conjunct $c_{j+1}^i$ exists if all of the following three conditions are satisfied: (1) the word just before the conjunct $c_j^i$ is a comma, (2) the word just before the comma has the same POS as the rightmost word of $c_j^i$, and (3) the Semantic Similarity between the two words referred to in the second condition is above a certain threshold. The third condition is used only for a pair of noun words.

Assume that the method has chosen “ryoin (both Houses)” as $c_1^i$ in Figure 1. Then, the method judges the existence of $c_2^i$ and concludes that it does not exist since the word just before $c_1^i$, “no (of),” is not a comma.

4.5 Substitution of Coordinate Structure

If a conjunct candidate includes an already identified coordinate structure, the Conjunct Similarity with an adjacent conjunct candidate is underestimated because of their different word lengths. To avoid this, the method substitutes the entire identified coordinate structure $crd^i$ with $c_0^i$.

In Figure 1, assuming that the method has identified $crd^i$, whose conjuncts are “ryoin (both Houses)” as $c_1^i$ and “ichiin (either House)” as $c_0^i$, the method substitutes “ryoin moshikuwa ichiin (both Houses or either House)” with “ichiin (either House)” and uses the new sentence “kokkai
no ichiin moshikuwa gikai no · · · (· · · of either House of the Diet or assemblies)” for the next process.

5 Neural Language Model

Our method uses neural language models (NLMs) to solve the problem of similarity calculation in the conventional method discussed in Section 4.3. In this section, we explain the basic concept of the NLM prior to giving a detailed explanation of our method.

An NLM is a language model working on a neural network (Bengio et al. 2003). It outputs a probability distribution of a word that follows an input word sequence. Figure 8 shows an example of an NLM constructed by an LSTM-based RNN (LSTM-NLM) (Sundermeyer, Schübler, and Ney 2010). The four hidden layers with recurrent connection in themselves can contain context of the input word sequence. We use this configuration of RNN in the evaluation experiment discussed in Section 7.

Using LSTM-NLMs, our method scores the following two presumptive characteristics of coordinate structure such that (1) any pair of conjuncts is similar, and (2) the fluency of a sentence is kept even if the conjuncts in a coordinate structure are swapped with each other. We assume that LSTM-NLMs will work well in our method because they theoretically store contexts of arbitrary length in recurrent connections and they abstract each input word sequence into a vector, which is called a distributed representation.

Fig. 8 Example of neural language model
6 Proposed Method

We believe that the conventional method has problems in identifying conjuncts because its approach is based on word alignment. Therefore, our method uses NLMs instead of word alignment in identifying conjuncts. On the other hand, we consider it reasonable to identify coordinate structures deterministically based on the descriptive rules unique to Japanese statutory sentences (Ishige 2004; Oshima 2005; Maeda 1983). Therefore, our method identifies coordinate structures and their conjuncts deterministically by the same flow as the conventional method. In this section, we explain our method’s procedure in each process in Figure 5, focusing on the points that differ from the conventional method.

6.1 Coordinator Extraction and Ranking

Table 2 shows the targeted coordinators and their priority in our method. Compared with the conventional method, our method adds two coordinators “kara (from)” and “sonotano (other2).” “kara (from)” is often used to indicate a range (e.g., a range of reference articles) (Maeda 1983). Judging from a syntactically tagged corpus of statutory sentences (Ogawa et al. 2011), we assign Priority 1 to “kara (from).” As described in Section 3.3, “sonotano (other2)” strictly has a different usage from “sonota (other1).”

Coordinate structures other than “kara (from)” and “sonotano (other2)” are formalized by Eq. (1). Coordinate structures formed by “kara (from)” and “sonotano (other2)” are formalized by Eqs. (2)\(^3\) and (3)\(^4\) respectively.

\[
\begin{align*}
crd_{kara}^i &= c_i^1 \cdot “kara” \cdot c_i^0 \cdot “made” , \quad (2) \\
crd_{sonotano}^i &= c_{n_i}^i \cdot “” \cdot c_{n_i-1}^i \cdot “” \cdot \ldots \cdot “” \cdot c_1^i \cdot “” \cdot “sonota” \cdot \ldots \cdot “no”. \quad (3)
\end{align*}
\]

| Priority | Coordinator                   |
|----------|-------------------------------|
| 1        | kara (から; from (A to B))    |
| 2        | oyobi (及び; and\(_L\))       |
|          | moshikuwa (若しくは; or\(_L\))|
| 3        | narabini (並びに; and\(_H\))  |
|          | matawa (又は; or\(_H\))      |
| 4        | sonota (その他; other\(_1\)) |
|          | sonotano (その他の; other\(_2\))|
|          | katsu (かつ; and)             |
|          | to (と; and)                  |
|          | ya (や; and)                  |

\(^3\) Equation (2) shows that if “kara (from)” constitutes a coordinate structure, it is always used with “made (to).”

\(^4\) As described in Section 3.3, “sonota” in “sonotano (other\(_2\))” becomes \(c_i^0\).
6.2 Candidate Extraction

In principle, our method extracts a candidate set $C_j^i$ of a conjunct $c_j^i$ in the same manner as the conventional method. However, we do not impose constraints from the thesaurus on our method because it does not use one, since it can cover the constraints using NLMs. Furthermore, our method introduces additional constraints to more completely cover the usage of coordinators in Japanese statutory sentences (Ishige 2004; Oshima 2005). The additional constraints are as follows:

- If $k_i$ is “oyobi (andL),” any candidate of $c_j^i$ should not include any coordinate structure with “oyobi (andL),” since the coordinator “oyobi (andL)” is used only for the lowest-layer coordinate structure.
- If $k_i$ is “matawa (orH),” any candidate of $c_j^i$ should not include any coordinate structure with “matawa (orH),” since the coordinator “matawa (orH)” is used only for the highest-layer coordinate structure.
- If $k_i$ is “matawa (orH),” any candidate of $c_j^i$ should not leave any coordinate structure with “moshikuwa (orL)” that can be included, since a coordinate structure with “moshikuwa (orL)” should be subordinate to a coordinate structure with “matawa (orH).”

6.3 Conjunct Identification

Our method identifies a conjunct based on two assumptions: (1) the conjuncts in any pair are alike and (2) the fluency of a sentence is maintained even if the conjuncts in a coordinate structure are swapped with each other. The first assumption has also been used in previous works for coordination analysis (Hara et al. 2009; Ficler and Goldberg 2016; Huddleston and Pullum 2002). Instead of our second assumption, some previous works (Ficler and Goldberg 2016; Huddleston and Pullum 2002) used another assumption, i.e., that a coordinate structure can be replaced by any of its conjuncts. However, this assumption on replacement may not work well for an incomplete coordinate structure, which our method generates while identifying coordinate structures with more than two conjuncts. Therefore, we adopt our second assumption because it works well even for an incomplete coordinate structure.

In our method, conjuncts $c_1^i$ and $c_0^i$ are identified simultaneously by Eq. (4), while conjuncts $c_j^i (j \geq 2)$ are identified by Eq. (5):

\[
(c_1^i, c_0^i) = \arg \max_{(c_l, c_r) \in C_1 \times C_0} \text{sim}(S, (c_l, c_r)) \times \text{flu}(S, (c_l, c_r)), \tag{4}
\]

\[
c_j^i = \arg \max_{c_i \in C_{j-1}^i} \text{sim}(S, (c_l, c_{j-1}^i)) \times \text{flu}(S, (c_l, c_{j-1}^i)), \tag{5}
\]
where $\text{sim}(S, (c_l, c_r))$ is a score of the similarity between two conjuncts $c_l$, $c_r$ (hereafter, **Similarity Score**) and $\text{flu}(S, (c_l, c_r))$ is a score of the fluency of the sentence $S$ after swapping two conjuncts $c_l$, $c_r$ (hereafter, **Fluency Score**). Referring to (Ficler and Goldberg 2016), our method calculates each score with two types of NLMs: (1) an **F-NLM** constructed by being fed words in forward order and (2) a **B-NLM** fed words in backward order. We use the two types of NLMs in order to use the context of the sentence even if a coordinate structure is at the beginning or ending of the sentence.

### 6.3.1 Similarity Score

The Similarity Score $\text{sim}(S, (c_l, c_r))$ is calculated by Eq. (6):

$$\text{sim}(S, (c_l, c_r)) = \text{sim}_f(S, (c_l, c_r)) + \text{sim}_b(S, (c_l, c_r)), \quad (6)$$

where

$$\text{sim}_f(S, (c_l, c_r)) = 1 + \text{sim}_c(\text{vec}_f(W_f), \text{vec}_f(W_{fr})), \quad (7)$$

$$\text{sim}_b(S, (c_l, c_r)) = 1 + \text{sim}_c(\text{vec}_b(W_{bl}), \text{vec}_b(W_{br})), \quad (8)$$

and $\text{sim}_c(u, v)$ is the cosine similarity between two vectors, $u$ and $v$. Adding one to $\text{sim}_c$ prevents $\text{sim}_f$ and $\text{sim}_b$ from becoming negative. $\text{vec}_f(W)$ denotes the output vector of the hidden layer nearest to the output layer after inputting a word sequence $W$ to the F-NLM, and $\text{vec}_b(W)$ denotes that nearest to the B-NLM. We expect that our method captures the similarity between word sequences in consideration of their contexts by calculating it based on the hidden layer’s output vectors.

The four word sequences $W_f$, $W_{fr}$, $W_{bl}$, $W_{br}$ are generated by Eqs. (9–12), respectively.

$$W_f = W_f \cdot c_l, \quad (9)$$

$$W_{fr} = W_f \cdot c_r, \quad (10)$$

$$W_{bl} = c_l \cdot W_b, \quad (11)$$

$$W_{br} = c_r \cdot W_b, \quad (12)$$

where $W_f$ represents the word sequence before $c_l$ in $S$ and $W_b$ represents the word sequence after $c_r$ in $S$. We assume that our method captures a broader context by using word sequences placed before and after the coordinate structure.
Assuming that we are calculating \( \text{sim}(S, \text{“ryoin”, “ichin”}) \) in Figure 1, the four word sequences \( W_{fl}, W_{fr}, W_{bl} \) and \( W_{br} \), respectively, become the following:

- \( W_{fl} = \text{“kokkai no ryoin_{ci}”} \)
- \( W_{fr} = \text{“kokkai no ichin_{ci}”} \)
- \( W_{bl} = \text{“ryoin_{ci} moshikuwa gikai no · · ·”} \)
- \( W_{br} = \text{“ichin_{ci} moshikuwa gikai no · · ·”} \)

### 6.3.2 Fluency Score

The Fluency Score \( \text{flu}(S, (c_l, c_r)) \) is calculated by Eq. (17):

\[
\text{flu}(S, (c_l, c_r)) = \text{flu}_f(W_s) + \text{flu}_b(W_s),
\]

where \( W_s \) is a word sequence after swapping two conjuncts \( c_l \) and \( c_r \) in \( S \). For example, \( W_s \) in calculating \( \text{flu}(S, \text{“ryoin”, “ichin”}) \) in Figure 1 is “kokkai no ichin moshikuwa ryoin moshikuwa gikai no · · ·” \( \text{flu}_f(W_s) \) and \( \text{flu}_b(W_s) \) denote the fluencies of \( W_s \) based on the F-NLM and B-NLM, respectively. These fluencies are calculated by the following two equations:

\[
\text{flu}_f(W_s) = \left( \prod_{t=1}^{[W_s]} P_f(w_t | w_1, w_2, \cdots, w_{t-1}) \right)^{\frac{1}{[W_s]}}
\]

\[
\text{flu}_b(W_s) = \left( \prod_{t=1}^{[W_s]} P_b(w_t | W_s[w_t], W_s[w_t-1], \cdots, w_{t+1}) \right)^{\frac{1}{[W_s]}}
\]

where \( P_f(w_t | w_1, w_2, \cdots, w_{t-1}) \) is the probability that \( w_t \) appears after the word sequence \( w_1, w_2, \cdots, w_{t-1} \) in the F-NLM. \( P_b(w_t | W_s[w_t], W_s[w_t-1], \cdots, w_{t+1}) \) is the corresponding probability in the B-NLM. We use the geometric mean to eliminate the effect of sentence length.

### 6.4 Further Conjunct Existence Judgment

Our method judges whether \( c_{j+1} \) exists in front of \( c_j \) in the same manner as the conventional method (see Section 4.4), except that we do not impose constraints based on Semantic Similarity on our method, since it does not use a thesaurus. In other words, our method judges that there is a further conjunct \( c_{j+1} \) next to the conjunct \( c_j \) if the following two conditions are satisfied:

- the word just before the conjunct \( c_j \) is a comma;
- the word just before the comma has the same POS as the rightmost word of \( c_j \).
However, using only these conditions sometimes leads to an incorrect judgment. For example, we assume that our method identifies a noun phrase conjunct “fusei kyoso no boshi (prevention of unfair competition)” in Figure 9. Consequently, our method mistakenly identifies “tame (in order to)” as another conjunct, since the word just before “fusei kyoso no boshi (prevention of unfair competition)” is a comma and the word before the comma, which is “tame (in order to),” is a noun.

Nevertheless, even if our method incorrectly judges the further conjunct’s existence, we do not consider imposing a constraint based on semantic similarity. This is because such a similarity-based constraint may reject words that are truly parts of conjuncts and lead to failure in identifying higher coordinate structures.

### 6.5 Substitution of Coordinate Structure

Our method substitutes $crd_i$ with $c_0^i$, in common with the conventional method.

### 7 Experiment

To evaluate the effectiveness of our method, we conducted an experiment on identifying coordinate structures in two Japanese acts (Administrative Procedure Act and Unfair Competition Prevention Act) contained in a syntactically tagged corpus of statutory sentences (Ogawa et al. 2011).

#### 7.1 Outline of Experiment

We constructed both input data and gold data from the above corpus (Ogawa et al. 2011). In the corpus, each sentence is annotated by hand with information on morphological analysis, bunsetsu segmentation, and dependency analysis. We constructed the input data by removing information on dependency analysis and extracting each parenthesized expression as another independent sentence. The size of the input data is 592 sentences with 792 coordinate structures.
We created the gold data using the dependency information in the corpus.

We built our NLMs using the LSTM-based RNNs (Sundermeyer et al. 2010) shown in Figure 8. Based on our preliminary experiment, we decided that each model has four hidden layers and that each layer consists of 650 LSTM units. A sequence of one-hot word vectors is fed into the model. In creating the one-hot word vector, we used basic forms of words obtained from a Japanese morphological analyzer MeCab (v0.98) (Kudo, Yamamoto, and Matsumoto 2004) with an IPA dictionary. We selected the 15,000 most frequent words for the vocabulary and added two special tokens, i.e., EOS (End Of Sentence) and UNK (UNKnown word), to the vocabulary. Thus, the numbers of input and output layer units are both 15,002.

The NLMs were trained using Chainer (v1.15.0).5 We used Softmax Cross Entropy as a loss function and updated the parameters by SGD (learning rate was 1) while clipping the L2 norm of gradient vectors to 5 and dropping out half of the randomly selected units. As the training data for our NLMs, we used 574,062 statutory sentences crawled from the Japanese Law Translation Database System (JLT)6 as of Nov. 2016.

For comparison, we implemented the conventional method (Matsuyama et al. 2012) on our own and conducted the same experiment. The parameters of the conventional method (e.g., threshold of the Semantic Similarity) were tuned using all of the input data.

In the evaluation, we measured recall, precision and F-measure for both conjuncts and coordinate structures. Precision $P$ is the ratio of correctly identified conjuncts or coordinate structures among all automatically identified ones. Recall $R$ is the ratio of correctly identified conjuncts or coordinate structures among all of those in the gold data. We judged that a conjunct has been correctly identified if its scope agreed completely with that of the gold data. Furthermore, we judged that a coordinate structure has been correctly identified if all of its conjuncts were correctly identified.

7.2 Experimental Results

Table 3 shows the experimental results of each method. Since our method was much superior to the conventional method in all evaluation indexes, we confirmed its effectiveness for coordination analysis of Japanese statutory sentences.

Figure 10 shows an example of a coordinate structure correctly identified by our method. The conventional method could not correctly identify the conjunct “shinsei wo shi yo to suru mono (persons planning to file applications).” As described in Section 4.3, this failure seems to be

5 http://chainer.org/
6 http://www.japaneselawtranslation.go.jp/
caused by a disparity of word length between the two conjuncts. On the other hand, our method transforms every conjunct candidate into a fixed-length vector by the NLMs, using context before and after the conjunct, and then directly calculates the score of a pair consisting of two conjunct candidates using the vectors. Therefore, we can conclude that our method correctly identified all of the conjuncts without being affected by their word lengths.

Next, Figure 11 shows an example of coordinate structures incorrectly identified by our method. In the gold data, “seifu (national governments)” and “chihokokyodantai (local governments)” are the conjuncts of the coordinate structure with “matawa (or $H$).” However, our method identified “gaikoku no seifu (foreign national governments)” as the first conjunct by mistake.

| Table 3  | Experimental results |
|----------|-----------------------|
|          | Our method            | Conventional method |
| Coordinate Structure | P 66.1% (463/700) | 46.8% (312/667) |
|          | R 64.6% (463/717)     | 43.5% (312/717)   |
|          | F 65.2                | 45.1              |
| Conjunct | P 83.0% (1,372/1,653) | 68.0% (1,019/1,499) |
|          | R 81.0% (1,372/1,694) | 60.2% (1,019/1,694) |
|          | F 82.0                | 63.8              |

Conventional method: ...

Our method (Gold data): ...

Fig. 11 Failed case in our method

【Any person who engages in public service for national or local foreign governments】
Article 18, paragraph (2), Item (i) of the Unfair Competition Prevention Act (Act No.47 of 1993)
8 Discussion

In this section, we reveal the characteristics and effectiveness of our method. First, we investigate our experimental results in more detail to evaluate our method as a coordination analysis method for Japanese statutory sentences. Second, we look further into the performance of the subtasks in our method.

8.1 Discussion of Experimental Results

The key characteristics of coordination in Japanese statutory sentences are the uniquely used coordinators and complex hierarchy. In this section, we look into our experimental results from the viewpoint of these characteristics.

8.1.1 Results by Coordinators

Our method introduces additional coordinators and rules to follow more closely the usage of coordinators in Japanese statutory sentences (Ishige 2004; Oshima 2005). We evaluated the performance of our method and that of the conventional method by the type of coordinators to compare the two methods’ characteristics more accurately.

Table 4 shows each method’s F-measures for coordinate structures sorted by type of coordinator. The second column indicates the frequency and its rate (among all coordinate structures) of coordinate structures sorted by type of coordinator. Our method dramatically outperformed the conventional method for all of the coordinators except “to (and).”

As described in Section 3, the four coordinators “matawa (orH),” “moshikuwa (orL),” “oyobi (andL),” and “narabini (andH)” constitute coordinate structures based on the descriptive rules.

| Coordinator          | Rate (Freq.) | Our method | Conventional method |
|----------------------|--------------|------------|---------------------|
| matawa (orH)         | 36.7% (263)  | 63.8       | 49.4                |
| moshikuwa (orL)      | 25.2% (181)  | 75.0       | 61.5                |
| oyobi (andL)         | 15.9% (114)  | 73.7       | 57.0                |
| narabini (andH)      | 3.5% (25)    | 54.9       | 23.5                |
| sonotano (other2)    | 6.8% (49)    | 57.1       | N/A                 |
| kara (from (A to B)) | 4.7% (32)    | 90.6       | N/A                 |
| only comma           | 3.1% (22)    | N/A        | N/A                 |
| sonota (other1)      | 1.8% (13)    | 30.8       | 0.0                 |
| katsu (and)          | 1.8% (13)    | 34.8       | 10.5                |
| to (and)             | 0.7% (5)     | 0.0        | 0.0                 |
unique to Japanese statutory sentences. Since 81.3% of all coordinate structures are composed of these four coordinators, the results for them had a huge effect on the total substantial improvement in Table 3. From this result, we confirmed the effectiveness of our method on hierarchical coordinate structure analysis for Japanese statutory sentences.

8.1.2 Results by Height of Coordinate Structure

Complex hierarchical coordinate structures appear frequently in Japanese statutory sentences, and it is not easy to completely identify their scope and hierarchy. Therefore, we evaluate how well our method can identify such hierarchical coordinate structures.

Table 5 shows each method’s F-measures sorted by height of coordinate structure. Since our method outperformed the conventional method at almost every height except the fifth one, we confirmed the effectiveness of our method for hierarchical coordinate structures frequently appearing in Japanese statutory sentences.

However, our method, as with the conventional method, tended to lose performance in higher coordinate structures. Three reasons seem to account for this result:

- Failure to identify lower coordinate structures affected the identification of the higher ones.
- The higher the coordinate structure, the longer its conjunct tends to become, since the structure also contains a lower coordinate structure. Therefore, the number of conjunct candidates tends to increase, and the probability of incorrect choice increases.
- When coordinators with the same priority appear in succession, there exists a hierarchical coordinate structure that our method cannot identify due to our rules on the analysis order of coordinators. We will delve into this in Section 8.2. Therefore, the difficulty of identifying a higher coordinate structure remains in our method.

| Height | Rate  | Freq. | Our method | Conventional method |
|--------|-------|-------|------------|---------------------|
| 1      | 71.4% | (512) | 74.1       | 54.3                |
| 2      | 21.2% | (152) | 50.2       | 17.9                |
| 3      | 5.4%  | (39)  | 26.2       | 22.2                |
| 4      | 1.5%  | (11)  | 8.7        | 0.0                 |
| 5      | 0.4%  | (3)   | 0.0        | 0.0                 |

Table 5 Results by layer (F-measure)
8.2 Discussion of Subtask Performance

Our method identifies **hierarchical coordinate structures** by sequentially executing the following two subtasks: (1) hierarchical structure identification and (2) coordinate structure identification. The hierarchical structure identification determines the hierarchy among the coordinate structures in a sentence. The coordinate structure identification determines the scopes of all conjuncts in a coordinate structure. In the following, we analyze the performance of each subtask.

8.2.1 Performance of Hierarchical Structure Identification

Our method roughly identifies a hierarchy among coordinate structures through coordinator ranking as described in Section 6.1, and then it identifies a complete hierarchy among them through coordinate structure identification as described in Sections 6.2–6.3. Since Japanese statutory sentences are written strictly according to Japanese legislation-drafting manuals, our coordinator ranking method is based on the heuristic rules obtained from the manuals, which consist of the Priority Rule and Position Rule defined in Section 4.1.

To evaluate our coordinator ranking method, we measured the percentage (hereafter, ranking accuracy) of the pairs consisting of two coordinators whose order relation in priority is correct. If two coordinate structures do not have hierarchy, we consider that any order between their coordinators is correct, since our method can correctly identify such coordinate structures regardless of their coordinator ranking. For example, if there are four coordinators in a sentence and two pairs have correct order relation in priority, then the percentage is 33.3% (= 2/4\(C_2\)).

The micro-averaged ranking accuracy for all of the 592 experimental input sentences in Section 7.1 was 95.7% (= 1,040/1,087). We can confirm that our heuristic coordinator ranking method can achieve high ranking accuracy and our heuristic rules are effective for coordinator ranking.

On the other hand, our coordinator ranking method did not succeed completely. Table 6 shows 47 pairs of coordinators whose order relation in priority were incorrectly identified by our method. Out of the 47 pairs, 29 pairs failed in identification because of the Priority Rule, while the other identification failures resulted from the Position Rule.

First, we discuss the cause of the failures from the Priority Rule shown in Table 2. In view of the Japanese legislation-drafting manuals referred to in Section 3, there is specific priority between two pairs of coordinators “matawa (or\(H\))” and “moshikuwa (or\(L\)),” and “oyobi (and\(L\))” and “narabini (and\(H\)).” However, there is no priority between other pairs of coordinators than the two pairs mentioned above. For example, there is no priority between “matawa (or\(H\))” and
“sonotano (other2).” This means that either a coordinate structure with “matawa (orH)” or that with “sonotano (other2)” can be the upper-layer one. Actually, the coordinate structure with “sonotano (other2)” becomes the upper-layer one in Sentence (a) of Figure 12, while that with “matawa (orH)” becomes the upper-layer one in Sentence (b) of Figure 12. The Priority Rule cannot capture the difference between such examples because the rule judges the priority only by the type of coordinator.

As another example, coordinate structures with “kara (from)” rarely have lower coordinate structures like that in Figure 13, and such a case does not appear in our test data (Ogawa et al.

| Larger coordinator   | Smaller coordinator | Rule violated | # |
|-----------------------|---------------------|---------------|---|
| oyobi (andL)          | to (and)            | Priority      | 2 |
|                       | matawa (orH)        | Priority      | 5 |
| moshikuwa (orL)       | sonotano (other2)   | Priority      | 1 |
|                       | moshikuwa (orL)     | Position      | 13|
|                       | matawa (orH)        | Priority      | 4 |
| narabini (andH)       | to (and)            | Priority      | 1 |
|                       | matawa (orH)        | Position      | 2 |
| matawa (orH)          | katsu (and)         | Priority      | 4 |
|                       | sonotano (other2)   | Priority      | 13|
| katsu (and)           | sonotano (other2)   | Position      | 1 |
| Total                 |                      |               | 47|

Fig. 12 Sentences having coordinate structures with “matawa (orH)” and “sonotano (other2)”: the coordinate structure with “sonotano (other2)” is upper-layer in Sentence (a), while that with “matawa (orH)” is upper-layer in Sentence (b).
Fig. 13  Sentence having a coordinate structure with “kara (from),” which has lower coordinate structures.

Fig. 14  Coordination analysis of a statutory sentence with multiple “moshikuwa (orL)”

2011). Our method cannot identify such hierarchical coordinate structures correctly since the priority of “kara (from)” is 1.

Next, we discuss the cause of the failures from the Position Rule. When plural coordinators have the same priority, our method regards a coordinator that comes earlier as the one that constitutes a lower-layer coordinate structure. However, there are some examples of hierarchical coordinate structures that violate this rule. In Figure 14, among four “moshikuwa (orL)” in this sentence, the second one constitutes the highest layer coordinate structure in the gold data. As described above, the Position Rule cannot assume that the second “moshikuwa (orL)” constitutes a higher layer coordinate structure than that of the third “moshikuwa (orL).”

Since the deterministic approach has limits for hierarchical structure identification, we need to adopt a more effective approach.

8.2.2 Performance of Coordinate Structure Identification

After identifying hierarchy among coordinate structures, our method deterministically identifies an internal structure of each coordinate structure, as described in Sections 6.2–6.5. In this section, we evaluate the performance of the coordinate structure identification from two points of view.

First, we conducted an experiment on coordination analysis under the condition of using an oracle coordinator ranking instead of the coordinator ranking generated in Section 6.1. We created the oracle from the hierarchy among coordinate structures acquired from the gold data.
We used the same experimental settings and evaluation measurement as Section 7.1, except for applying the oracle.

Table 7 shows the experimental results. Even if the hierarchy among coordinate structures is given, the F-measure of coordination analysis was less than 70 points. We can confirm that coordinate structure identification is an inherently difficult subtask. On the other hand, coordination analysis with oracle coordinator ranking identified 29 more coordinate structures correctly and achieved 4.7 points higher performance at F-measure than the analysis without oracle. Therefore, we can confirm that it is fruitful to improve the coordinator ranking method.

Second, we conducted an ablation study to evaluate the contributions of the Similarity Score (see Section 6.3.1) and the Fluency Score (see Section 6.3.2) used for conjunct identification. We implemented two experimental methods that (1) identifies each conjunct only with the Similarity Score and (2) identifies each conjunct only with the Fluency Score. Then we conducted an experiment on coordination analysis using the two methods described above and using our proposed method that uses both the Similarity Score and the Fluency Score. We used the same experimental settings and evaluation measurement as Section 7.1.

Table 8 shows the experimental results. Our proposed method using both functions achieved
the best performance in every evaluation measurement. From this result, we can confirm that it is effective to use both the Similarity Score and the Fluency Score.

9 Conclusion

In this paper, we proposed a hierarchical coordinate structure analysis method for Japanese statutory sentences using neural language models. Our method deterministically identifies the scope of all conjunctions in coordinate structures, using NLMs to score each candidate. The deterministic algorithm of our method is based on well-established descriptive style rules of a Japanese legislation-drafting manual. The experimental results show that our method was much more effective than the conventional method.

In future work, we aim to improve performance by resolving the problem of hierarchy identification and by refining the NLMs and scoring function.

Acknowledgement

This research was partly supported by the Japan Society for the Promotion of Science KAKENHI Grant-in-Aid for Scientific Research (A) No. 26240050 and (C) No. 16K00300.

Reference

Agarwal, R. and Boggess, L. (1992). “A Simple But Useful Approach to Conjunct Identification.” In Proceedings of the 30th Annual Meeting of the Association for Computational Linguistics, pp. 15–21.

Bengio, Y., Ducharme, R., Vincent, P., and Jauvin, C. (2003). “A Neural Probabilistic Language Model.” Journal of Machine Learning Research, 3, pp. 1137–1155.

Chantree, F., Kilgarriff, A., de Roeck, A., and Willis, A. (2005). “Disambiguating Coordinations using Word Distribution Information.” In Proceedings of the International Conference on Recent Advances in Natural Language Processing, pp. 144–151.

Dubey, A., Keller, F., and Sturt, P. (2006). “Integrating Syntactic Priming into an Incremental Probabilistic Parser, with an Application to Psycholinguistic Modeling.” In Proceedings of the Joint 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics, pp. 417–424.
Ficler, J. and Goldberg, Y. (2016). “A Neural Network for Coordination Boundary Prediction.” In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pp. 23–32.

Goldberg, M. (1999). “An Unsupervised Model for Statistically Determining Coordinate Phrase Attachment.” In Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics, pp. 610–614.

Hagiwara, M., Ogawa, Y., and Toyama, K. (2009). “Bootstrapping-based extraction of dictionary terms from unsegmented legal text.” In Hattori, H., Kawamura, T., Idé, T., Yokoo, M., and Murakami, Y. (Eds.), New Frontiers in Artificial Intelligence: JSAI 2008 Conference and Workshops, Revised Selected Papers, Lecture Notes in Computer Science, Vol. 5447, pp. 213–227. Springer.

Hanamoto, A., Matsuzaki, T., and Tsujii, J. (2012). “Coordination structure analysis using dual decomposition.” In Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics, pp. 430–438.

Hara, K., Shimbo, M., Okuma, H., and Matsumoto, Y. (2009). “Coordinate Structure Analysis with Global Structural Constraints and Alignment-based Local Features.” In Proceedings of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing, pp. 967–975.

Huddleston, R. D. and Pullum, G. K. (2002). The Cambridge Grammar of the English Language. Cambridge University Press.

Ishige, M. (2004). Jichirippojitsumu no tame no hoseishitsumu shokai (4th edition). Gyosei. (In Japanese).

Kawahara, D. and Kurohashi, S. (2006). “A Fully-lexicalized Probabilistic Model for Japanese Syntactic and Case Structure Analysis.” In Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics, pp. 176–183.

Kawahara, D. and Kurohashi, S. (2008). “Coordination Disambiguation without Any Similarities.” In Proceedings of the 22nd International Conference on Computational Linguistics, pp. 425–432.

Kudo, T., Yamamoto, K., and Matsumoto, Y. (2004). “Applying Conditional Random Fields to Japanese Morphological Analysis.” In Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, pp. 230–237.

Kurohashi, S. and Nagao, M. (1994). “A Syntactic Analysis Method of Long Japanese Sentences Based on the Detection of Conjunctive Structures.” Computational Linguistics, 20 (4),
pp. 507–534.

Maeda, M. (1983). *Workbook Houseishitsumu (completely revised)*. Gyosei. (In Japanese).

Matsuyama, H., Shirai, K., and Shimazu, A. (2012). “Horei bunsho wo taisho ni shita heiretsukozo kaiseki.” In *Proceedings of the 18th Annual Meeting of the Association for Natural Language Processing*, pp. 975–978. (In Japanese).

Ogawa, Y., Yamada, M., Kato, R., and Toyama, K. (2011). “Design and Compilation of Syntactically Tagged Corpus of Japanese Statutory Sentences.” In *New Frontiers in Artificial Intelligence: JSAI 2010 Conference and Workshops, Revised Selected Papers, Lecture Notes in Computer Science*, Vol. 6797, pp. 141–152. Springer.

Oshima, T. (2005). *Hoseishitsumu no kiso chishiki (3rd edition)*. Dai-ichi Hoki. (In Japanese).

Resnik, P. (1999). “Semantic Similarity in a Taxonomy: An Information-based Measure and Its Application to Problems of Ambiguity in Natural Language.” *Journal of Artificial Intelligence Research*, 11, pp. 95–130.

Roh, Y.-H., Lee, K.-Y., Choi, S.-K., Kwon, O.-W., and Kim, Y.-G. (2008). “Recognizing Coordinate Structures for Machine Translation of English Patent Documents.” In *Proceedings of the 22nd Pacific Asia Conference on Language, Information and Computation*, pp. 460–466.

Shirai, S., Ooyama, Y., Ikehara, S., Miyazaki, M., and Yokoo, A. (1998). “Introduction to “Goi - Taikei: A Japanese Lexicon.” Tech. rep. 106 (1998-IM-034), ATR Spoken Language Translation Research Laboratories and NTT Communication Science Laboratories, NTT Communication Science Laboratories, Faculty of Engineering Tottori University, Faculty of Engineering Niigata University, ATR Spoken Language Translation Research Laboratories. (In Japanese).

Sundermeyer, M., Schullüter, R., and Ney, H. (2010). “LSTM Neural Networks for Language Modeling.” In *Proceedings of the INTERSPEECH 2010*, pp. 194–197.

Yamada, D. and Shimazu, A. (2006). “Horeibun no gengoteki tokucho wo riyoshita kadokusei kojo no tame no hyoji.” In *Proceedings of the 12th Annual Meeting of the Association for Natural Language Processing*, pp. 196–199. (In Japanese).

**Takahiro Yamakoshi**: received his Master’s degree in Information Science from Nagoya University in 2018. He is currently a Ph.D. student at Graduate School of Informatics, Nagoya University. His research interests include natural language processing, especially document creation support. He is a member of the ANLP and the IPSJ.
Tomohiro Ohno: received his Ph.D. degree in Information Science from Nagoya University in 2007. He is currently an associate professor at Graduate School of Advanced Science and Technology, Tokyo Denki University. His research interests include natural language processing and spoken language processing. He is a member of the IPSJ and the IEICE.

Yasuhiro Ogawa: received his Ph.D. degree in Engineering from Nagoya University in 2000. He is currently an associate professor at the Information Technology Center at Nagoya University and the Graduate School of Informatics, Nagoya University. His research interest is natural language processing including machine translation, legal informatics and e-legislation.

Makoto Nakamura: received his Ph.D. degree in Information Science from JAIST in 2004. After being an assistant professor at JAIST and a designated assistant and associate professor at Japan Legal Information Institute, Graduate School of Law, Nagoya University, he is currently an associate professor at Department of Engineering, Niigata Institute of Technology. His research interests include legal text processing and simulation of language evolution.

Katsuhiko Toyama: received his Ph.D. degree in Engineering in 1989. He is currently a professor in the Information Technology Center and the Graduate School of Informatics at Nagoya University. His current research interests include natural language processing, especially legal text processing, and legislation support systems. He is a member of the IPSJ, the JSAI, and the IEICE.

(Received December 21, 2017)
(Revised March 13, 2018)
(Accepted April 27, 2018)