# Japanese Word Reordering Based on Topological Sort

Peng Sun[1], Tomohiro Ohno[2][a] and Shigeki Matsubara[1,3][b]

[1]*Graduate School of Informatics, Nagoya University, Nagoya, Japan*

[2]*Graduate School of Advanced Science and Technology, Tokyo Denki University, Tokyo, Japan*

[3]*Information & Communications, Nagoya University, Nagoya, Japan*

Keywords: Writing Assistance, Elaboration Support, Paraphrasing, Readability, Sorting Algorithm.

Abstract: In Japanese, some sentences are grammatically well-formed, but not easy to read. This paper proposes a method for Japanese word reordering, which first adapts a Japanese BERT model to predict suitable pairwise orderings between two words in a sentence, and then converts the predicted results into a graph. The vertices in the graph represent the words in the sentence. The edges represent the pairwise orderings between two words. Finally, topological sort is applied to find the correct word ordering in a sentence by visiting each graph vertex. We conducted an evaluation experiment with uneasy-to-read Japanese sentences created from newspaper article sentences.

## 1 INTRODUCTION

The Japanese language has a relatively free word order, making it possible to write meaningful sentences without paying much attention to the word order. In practice, however, some sort of preference in the Japanese word order exists. These preferences may lead to the generation of some grammatically correct, but uneasy-to-read sentences.

Word reordering is the arrangement of words in an input sentence such that the word order can easily be read. It has been studied as a basic technique for applications like writing assistance and sentence generation. Several methods on Japanese word reordering have been proposed (Uchimoto et al., 2000; Yokobayashi et al., 2004; Ohno et al., 2015; Miyachi et al., 2021). All of them are based on the dependency relations between words, which are assumed to be given as the input by preliminarily executing dependency parsing or partially given by concurrently executing dependency parsing. This is considering the close relationship between the word order and dependency relations. That is, the possible word order is restricted by dependency relations. However, if the word order of an input sentence cannot be easily read, the accuracy of word reordering decreases under the influence of the accuracy of dependency parsing, which tends to decrease.

This paper proposes a Japanese word reordering method without dependency parsing. The strategy does not use dependency information; thus, it cannot be affected by dependency parsing errors, which is an advantage. By contrast, word reordering without dependency information causes another problem. That is, the possible permutation of words in an input sentence may not be narrowed down based on the dependency information; thus, the computational cost explosively increases. In this work, we use topological sort (Kahn, 1962; Tarjan, 1976) to resolve the above mentioned problem. Topological sort is an algorithm for the linear ordering of the vertices of a directed acyclic graph (DAG) to efficiently find the appropriate word order in a sentence.

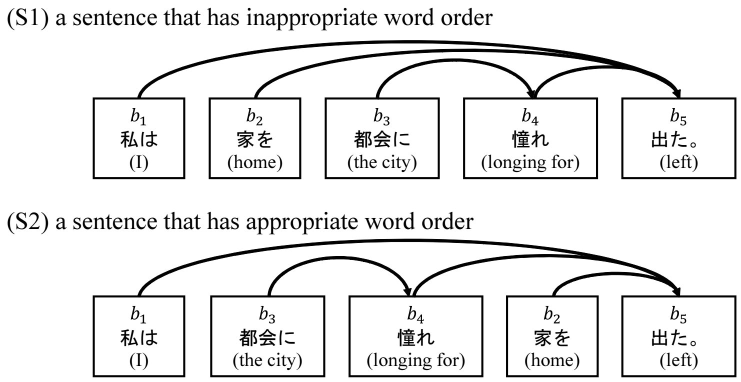
Our method first predicts the suitable pairwise orderings between two words in a sentence using a Japanese BERT model. It then converts the predicted results into a graph, where a vertex and an edge represent a word in the sentence and a pairwise ordering between two words, respectively. Finally, it identifies the appropriate word order by applying topological sort to the graph. We conducted an evaluation experiment and confirmed the effectiveness of our method on Japanese word reordering.

The remainder of this paper is organized as follows: Section 2 explains the Japanese word reordering; Section3 introduces topological sort and its application to natural language processing; Section 4

[a] https://orcid.org/0000-0001-7015-7714
[b] https://orcid.org/0000-0003-0416-3635



Sun, P., Ohno, T. and Matsubara, S.
Japanese Word Reordering Based on Topological Sort.
DOI: 10.5220/0011775000003393
In *Proceedings of the 15th International Conference on Agents and Artificial Intelligence (ICAART 2023) - Volume 3*, pages 768-775
ISBN: 978-989-758-623-1; ISSN: 2184-433X




The sentence (S1) and (S2) have the same meaning that is translated as "I left home longing for the city" in English.
The difference between (S1) and (S2) is just their word orders in Japanese.

Figure 1: Examples of inappropriate and appropriate word orders in Japanese.

presents the proposed method of Japanese word reordering based on topological sort; Section 5 explains the conducted experiment for quantitatively evaluating the proposed method; and Section 6 summarizes this work.

## 2 JAPANESE WORD ORDER AND DEPENDENCY

We propose herein a word reordering method aimed for the elaboration support of Japanese natives. We assumed that the input sentences were not grammatically incorrect, but had an uneasy-to-read word order. The reasons behind this assumption are as follows: Even Japanese natives often create such sentences unless they pay attention to the Japanese word order preference, and it is not difficult for Japanese natives to create grammatically correct sentences due to the relatively free word order.

Many studies have been conducted on the Japanese word order preference in linguistics (e.g., (Uchimoto et al., 2000; Kuribayashi et al., 2020; Fujihara et al., 2022)). Case elements have been reported to be basically placed in the order of a nominative, a dative, and an accusative, and that the basic order of case elements is often changed by influence from other factors, such as long dependencies.

Figure 1 shows two Japanese sentences, which differ only in their word order, are grammatically correct, and have the same meaning (i.e., translated as "I left home longing for the city."). The box and the arrow in the figure express a *bunsetsu*[1] and a dependency relation, respectively. The word order of S1 is more difficult to read than that of S2 because the distance between the bunsetsu "家を (home)" and its modified bunsetsu "出た (left)" is large, causing the loads on the working memory to become large (Yoshida et al., 2014). This suggests that the dependency information is useful in reordering words such that the word order becomes easier to read.

If a Japanese sentence is grammatically correct, the sentence satisfies the following constraints on the word order and dependency:

1. No dependency is directed from right to left.
2. Dependencies do not cross each other.
3. Each bunsetsu, except for the final one in a sentence, depends on only one bunsetsu.

Therefore, if the dependencies between all words in a sentence can be parsed before the word reordering, we can narrow down the candidates of the most appropriate word order according to the dependencies. In Figure 1, the possible word order can be narrowed down to six permutations (i.e., "$b_1b_2b_3b_4b_5$", "$b_1b_3b_4b_2b_5$", "$b_2b_1b_3b_4b_5$", "$b_2b_3b_4b_1b_5$", "$b_3b_4b_1b_2b_5$" and "$b_3b_4b_2b_1b_5$") due to the dependency relations. Most of the previous methods on word reordering (Belz and Kow, 2011; Filippova and Strube, 2007; Harbusch et al., 2006; Kruijff et al., 2001; Ringger et al., 2004; Shaw and Hatzivassiloglou, 1999; Yokobayashi et al., 2004) premise that dependency parsing is preliminarily performed and identify the most appropriate word order among the possible word order candidates that satisfy the above mentioned constraints using the preference

[1] *Bunsetsu* is a Japanese linguistic unit that roughly corresponds to a basic phrase in English. A bunsetsu consists of one independent word and zero or more ancillary words.

A dependency relation in Japanese is a modification relation, in which a modifier bunsetsu depends on a modified bunsetsu. That is, the modifier and modified bunsetsus work as the modifier and that which is modified, respectively.

between word order and dependencies. However, the word order of S1 is thought to be more difficult to parse than that of S2 because dependency parsers are usually trained on syntactically annotated corpora, in which sentences have an appropriate word order, such as that in S2.

The above mentioned constraints between the word order and the dependencies cannot be used in word reordering if we choose not to use dependency parsing in word reordering. Thus, it is necessary to select the best permutation among all possible word permutations in an input sentence. If an input sentence has $n$ bunsetsus, the number of all possible permutations is $n!$, which in Figure 1 is $5! = 120$. Therefore, an efficient algorithm must be employed to explore the best permutation among the $n!$ candidates. It cannot use the dependency information, which is useful in word reordering. Hence, a model that can predict the appropriate word order without directly using the dependency information must be employed.

# 3 TOPOLOGICAL SORT

Topological sort (Kahn, 1962; Tarjan, 1976) is an algorithm that linearly arranges all the vertices of a DAG according to the edge directions. All vertices are briefly linearly ordered, such that every vertex precedes its next neighbor vertices connected by the outgoing edge from the vertex. That is, for every edge $v \rightarrow u$ directed from a vertex $v$ to a vertex $u$ in a DAG, the algorithm decides the order, such that $v$ comes before $u$.

Topological sort is used for some natural language processing applications that sort something. Prabhumoye et al. (2000) proposed a method for identifying the most appropriate sentence order in a document using topological sort (Prabhumoye et al., 2020; Keswani and Jhamtani, 2021). Sentence ordering is the task of arranging all sentences in a given document, such that the document consistency is maximized and applied (e.g., multi-document summarization (Barzilay and Elhadad, 2002; Nallapati et al., 2017), and cooking procedure generation (Nallapati et al., 2017)). In their work, Prabhumoye et al. predicted each relative ordering between two sentences in a document. Each predicted result was then regarded as a constraint of the anteroposterior relation of a sentence pair. A DAG was made to express a set of constraints. They found the most appropriate order of sentences by applying topological sort to the DAG. The word reordering task is similar to the sentence ordering task in terms of sorting elements in a unit; hence, topological sort is thought to be applicable to the word reordering task in the same way as in the sentence ordering task.

The time complexity of topological sort is $O(|V|+|E|)$, where $V$ and $E$ are a set of all the vertices and edges in a DAG, respectively. If we use topological sort for word reordering of an input Japanese sentence, a DAG has a vertex and an edge expressing a bunsetsu in the sentence and a relative ordering between two bunsetsus, respectively. In this case, if an input sentence has $n$ bunsetsus, because $|V|$ is $n$ and $|E|$ is the combination ${}_nC_2 = \binom{n}{2} = n*(n-1)/2$, the time complexity of the topological sort is $O(n^2)$. Therefore, topological sort is expected to resolve the high computational cost problem described in Section 2.

# 4 PROPOSED METHOD

In our method, a grammatically correct, but uneasy-to-read sentence is assumed as the input. Our method reorders all bunsetsus in the input sentence such that the reordered sentence becomes easy-to-read. Figure 2 shows the method framework.

First, a sequence of bunsetsus of a sentence, which has an uneasy-to-read word order, is input. In Figure 1(S1), the input sentence is “私は (I) / 家を (home) / 都会に (the city) / 憧れ (longing for) / 出た。(left.)”. Our method extracts every combination of any two bunsetsus among a set of all bunsetsus, excluding the final one in an input sentence (In Figure 1, {“私は (I),” “家を (home),” “都会に (the city),” “憧れ (longing for)”}). It then predicts the relative ordering of each pair, that is, the anteroposterior relation between two bunsetsus, using a Japanese BERT model. If both an input sentence and its reordered sentence are assumed to be grammatically correct, the final bunsetsu of the input sentence always becomes the final bunsetsu of the output sentence without depending on the word reordering. This is why the final bunsetsu is excluded.

Second, a set of predicted anteroposterior relations is converted into a directed graph, where a vertex and an edge represent a word in the sentence and a predicted anteroposterior relation between two bunsetsus, respectively. If the created graph is a directed cyclic graph (DCG), it will be converted into a DAG, to which topological sort can be applied. The conversion method is based on the source code of the method proposed by Prabhumoye et al. (2000). In the topological sort process during conversion, each time a closed path is found, the last edge of the closed path (i.e., the edge that returns to a visited vertex) is deleted until no more closed paths are left.

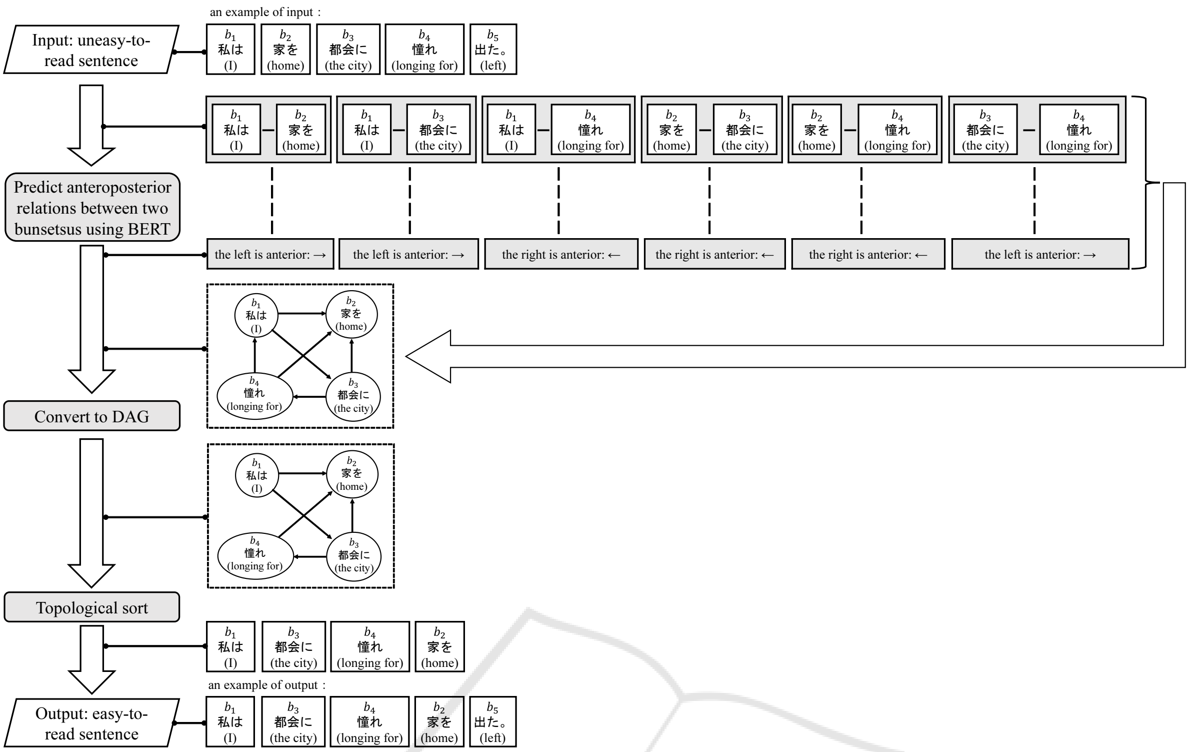



Figure 2: Framework of our method.

Finally, topological sort is applied to the DAG created above. Each vertex (each bunsetsu) is then linearly ordered. Note that our method uses topological sort based on the depth-first search (Tarjan, 1976).

## 4.1 Model

Figure 3 shows an overview of our BERT model that predicts the anteroposterior relation between two bunsetsus. Here, $B = b_1 b_2 \cdots b_n$ expresses a sequence of bunsetsus in an input sentence, while $b_i,\ b_j\,(1 \leq i < j \leq n-1)$ expresses the two bunsetsus, between which the anteroposterior relation is predicted by our BERT model. During this time, the input to BERT is the concatenation "[CLS] $b_i$ [SEP] $b_j$ [SEP] $b_1 b_2 \cdots b_n$ [SEP]", where subword division is performed.

An input sentence, even if not easily readable, has a partially appropriate word order because the input sentence is grammatically correct. The relative location of many bunsetsus after reordering tends to be maintained in the same order as that of the input sentence. In addition, an input sentence contains the dependency information between bunsetsus, which is considered to be effective for predicting the relative ordering between two bunsetsus as described in Section 2. These are the reasons why our method takes the entire input sentence $b_1 b_2 \cdots b_n$ into the BERT model.

The BERT model outputs a bidimensional probability distribution that expresses how appropriate the anteroposterior relation (i.e., either $b_i$ or $b_j$ precedes the other) is in readability. The anteroposterior relation with a higher probability becomes the predicted result used as the constraint on the relative ordering between the two bunsetus. The prediction by our BERT model is performed for each combination of any two bunsetsus created among $\{b_1,\, b_2,\, \cdots,\, b_{n-1}\}$.

## 4.2 Creation of Training Data

The kind of data to use as the training data for our BERT model is also important. As a simple approach, it can use the anteroposterior relations included in easy-to-read sentences (e.g. newspaper article sentences) as the training data. This approach makes two training events from one anteroposterior relation between two bunsetsus $b_i$ and $b_j (i < j)$ in a sentence. One is "$b_i \rightarrow b_j$" labeled with "the left is anterior", and the other is "$b_i \leftarrow b_j$" labeled with "the right is anterior". The number of labels that is "the left is anterior" becomes equal to that of "the right is anterior" in these training data. However, it can be thought that no such tendency in humans actually creates uneasy-to-read sentences.

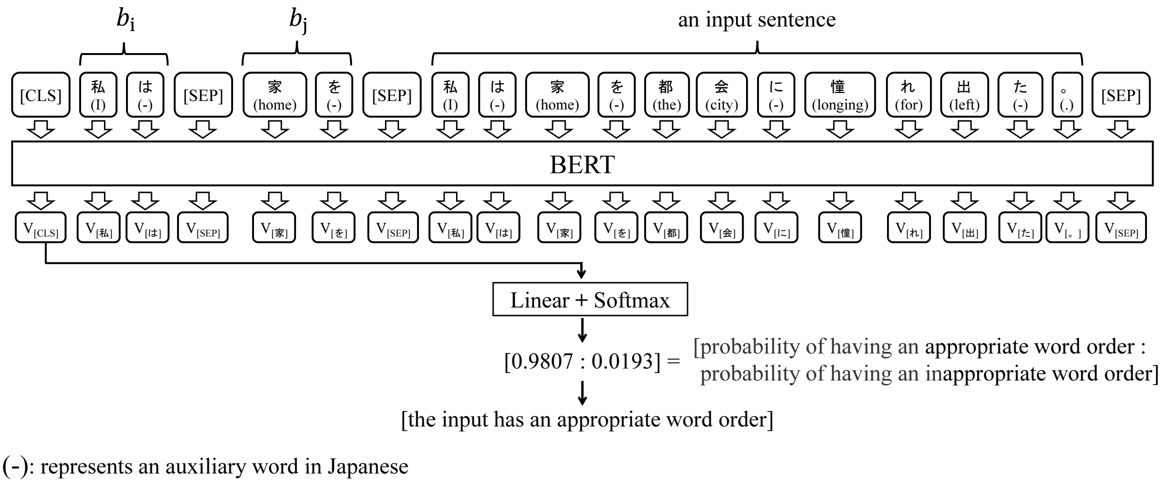



Figure 3: Example of predicting the anteroposterior relation between two bunsetsus.

To train the tendency of humans to actually create uneasy-to-read sentences, we must annotate the combination of two bunsetsus in sentences with a label. However, annotation costs are very high.

To solve this problem, based on the assumption that newspaper article sentences are written in an easy-to-read word order, we mechanically create a pseudo-sentence with an uneasy-to-read word order from each newspaper article sentence following the procedure below:

1. Remove commas from the newspaper article sentences.
2. Find a bunsetsu modified by multiple bunsetsus from the sentence end.
3. Randomly change the order of sub-trees that modify this bunsetsu.
4. Iterate 2 and 3 until you reach the beginning of the sentence.

The created pseudo-sentences satisfy the three constraints on word order and dependency described in Section 2; thus, they are assumed to be grammatically correct, but have an uneasy-to-read word order different from that of newspaper article sentences.

With the above mentioned procedure, we can create many training data with no cost if a corpus exists, including newspaper article sentences annotated with information on morphological analysis, bunsetsu segmentation, and dependency analysis.

# 5 EXPERIMENT

We conducted an experiment on the word reordering of uneasy-to-read sentences using Japanese newspaper articles to evaluate the effectiveness of our method.

## 5.1 Experiment Outline

For the experiment, 1000 sentences with an uneasy-to-read word order artificially created from newspaper articles sentences in the Kyoto Text Corpus were used as the test and development data (Kuroashi and Nagao, 1998), respectively, which is annotated with information on morphological analysis, bunsetsu segmentation, and dependency analysis. The creation was based on the following procedure:

(1) Mechanically create some new sentences from a newspaper article sentence based on the procedure described in Section 4.2. All the sentences created by this step have a different word order from the original newspaper article sentence but maintain the same dependency relations.

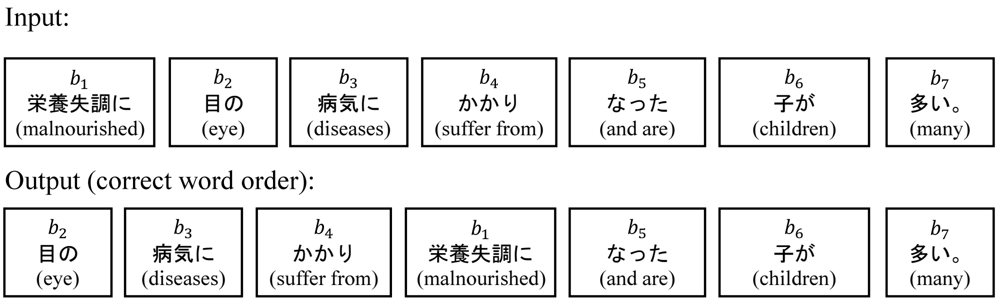
(2) Among the above mentioned sentences, manually select one sentence considered as sometimes written by Japanese native speakers.

(3) Manually add commas to make it as easy as possible to read in that word order.

We used 34,199 sentences created by the procedure described in Section 4.2 as the training data. Among which, 27,263 sentences have a different order from the original newspaper article sentences, and 6936 sentences have the same order by chance. These 34,199 sentences had a more-than-equal-to 3 bunsetsus because extremely short original sentences, whose word order can be decided based on the constraints described in Section 2 without using BERT, were deleted. Note that every original newspaper article

Table 1: Experimental Results.

| | pair agreement | sentence agreement |
|---|---|---|
| our method | 88.49% (28,105/31,760) | 40.60% (406/1,000) |
| [random] | 50.59% (16,070/31,760) | 4.60% (46/1,000) |
| [no reordering] | 75.48% (23,973/31,760) | 0.00% (0/1,000) |




The sentence is translated as "Many children suffer from eye diseases and are malnourished." in English.

Figure 4: Example of sentences correctly reordered by our method.

sentence of the training data is different from that of the test and development data.

In the evaluation, we obtained two the following measurements are defined by Uchimoto et al. (2000) and Miyachi et al. (2021):

- Sentence agreement: percentage of the output sentences in which the word order entirely agrees with that of the original sentence.
- Pair agreement: percentage of the pairs of bunsetsus, whose word order agrees with the word order in the original sentence.

The two following baselines were established for comparison:

- [random]: randomly reorders all bunsetsus except the final bunsetsu in an input sentence.
- [no reordering]: outputs an input sentence without changing the word order.

## 5.2 Experimental Results

Table 1 shows the experimental results on the word reordering of our method and each baseline. Our method dramatically outperformed the two baselines in both the pair and sentence agreements.

Figure 4 illustrates an example of the sentences, in which all bunsetsus were correctly reordered by our method. In this example, our method correctly moved a bunsetsu "栄養失調に (malnourished)" just before the dependent bunsetsu "なった (and are)." The distance of the two bunsetsus, where a dependency relation existed, in the output sentence became shorter than that of the input one. Moreover, the readability of the output sentence was improved. These results confirmed the effectiveness of our method on Japanese word reordering.

In Section 4, the graph became either a DCG, which was then converted to a DAG, or a DAG when a directed graph was created from the anteroposterior relations predicted by BERT. We re-measured the pair and sentence agreements separately for sentences became DCGs and DAGs. Tables 2 and 3 show the pair and sentence agreements of our method and [no reordering], respectively. The results of [no reordering] (i.e., input sentences) revealed not much of a difference between the DAGs and the DCGs. In contrast, the results of our method showed that both agreements for sentences that became DAGs were much higher than those for sentences that became DCGs. Some sentences were converted into a DCG because the prediction by BERT contained some incorrect anteroposterior relations. Thus, even if the DCGs were converted into DAGs, the agreements were thought to have decreased because all incorrect edges in the graphs were not always deleted. Resolving the agreement decrease required the improvement of the prediction accuracy of the anteroposterior relations by BERT and the conversion of a DCG to a DAG by the proper removal of incorrect edges. These issues are to be addressed in the future.

# 6 CONCLUSIONS

This paper proposed a method for the word reordering of uneasy-to-read sentences without executing dependency parsing. The method used BERT to predict the

Table 2: Pair agreement separately for sentences that had become DCGs and sentences that had become DAGs.

| | DAG | DCG |
|---|---|---|
| our method | 91.51% (14,841/16,218) | 85.34% (13,264/15,542) |
| [no reordering] | 75.74% (12,284/16,218) | 75.21% (11,689/15,542) |

Table 3: Sentence agreement separately for sentences that had become DCGs and sentences that had become DAGs.

| | DAG | DCG |
|---|---|---|
| our method | 50.29% (350/696) | 18.42% (56/304) |
| [no reordering] | 0.00% (0/696) | 0.00% (0/304) |

anteroposterior relations between two bunsetsus and applied topological sort to the predicted results. The effectiveness of our method was confirmed through the evaluation experiments using sentences with an uneasy-to-read word order created from newspaper article sentences.

For future works, we would like to improve the pair and sentence agreements on word reordering. We also intend to build an uneasy-to-read sentence corpus that humans actually create.

## ACKNOWLEDGEMENTS

This work was partially supported by JSPS KAKENHI Grand Number JP19K12127.

## REFERENCES

Barzilay, R. and Elhadad, N. (2002). Inferring strategies for sentence ordering in multidocument news summarization. *Journal of Artificial Intelligence Research*, 17(1):35–55.

Belz, A. and Kow, E. (2011). Discrete vs. continuous rating scales for language evaluation in nlp. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 230–235.

Filippova, K. and Strube, M. (2007). Generating constituent order in german clauses. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 320–327.

Fujihara, R., Kuribayashi, T., Abe, K., Tokuhisa, R., and Inui, K. (2022). Topicalization in language models: A case study on Japanese. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 851–862.

Harbusch, K., Kempen, G., Van Breugel, C., and Koch, U. (2006). A generation-oriented workbench for performance grammar: Capturing linear order variability in german and dutch. In *Proceedings of the 4th International Natural Language Generation Conference*, pages 9–11.

Kahn, A. B. (1962). Topological sorting of large networks. *Communications of the ACM*, 5(11):558–562.

Keswani, V. and Jhamtani, H. (2021). Formulating neural sentence ordering as the asymmetric traveling salesman problem. In *Proceedings of the 14th International Conference on Natural Language Generation*, pages 128–139.

Kruijff, G.-J. M., Kruijff-Korbayová, I., Bateman, J., and Teich, E. (2001). Linear order as higher-level decision: Information structure in strategic and tactical generation. In *Proceedings of the 8th European Workshop on Natural Language Generation*, pages 74–83.

Kuribayashi, T., Ito, T., Suzuki, J., and Inui, K. (2020). Language models as an alternative evaluator of word order hypotheses: A case study in Japanese. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 488–504.

Kuroashi, S. and Nagao, M. (1998). Building a Japanese parsed corpus while improving the parsing system. In *Proceedings of the 1st International Conference on Language Resources and Evaluation*, pages 719–724.

Miyachi, K., Ohno, T., and Matsubara, S. (2021). Word reordering and comma insertion integrated with shift-reduce dependency parsing. In *Proceedings of the 13th International Conference on Agents and Artificial Intelligence - Volume 2*, pages 1144–1150.

Nallapati, R., Zhai, F., and Zhou, B. (2017). Summarunner: A recurrent neural network based sequence model for extractive summarization of documents. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence*, pages 3075–3081.

Ohno, T., Yoshida, K., Kato, K., and Matsubara, S. (2015). Japanese word reordering executed concurrently with dependency parsing and its evaluation. In *Proceedings of the 15th European Workshop on Natural Language Generation*, pages 61–65.

Prabhumoye, S., Salakhutdinov, R., and Black, A. W. (2020). Topological sort for sentence ordering. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 2783–2792.

Ringger, E., Gamon, M., Moore, R. C., Rojas, D. M., Smets, M., and Corston-Oliver, S. (2004). Linguistically informed statistical models of constituent structure for ordering in sentence realization. In *Proceedings of the 20th International Conference on Computational Linguistics*, pages 673–679.

Shaw, J. and Hatzivassiloglou, V. (1999). Ordering among premodifiers. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, pages 135–143.

Tarjan, R. E. (1976). Edge-disjoint spanning trees and depth-first search. *Acta Informatica*, 6(2):171–185.

Uchimoto, K., Murata, M., Ma, Q., Sekine, S., and Isahara, H. (2000). Word order acquisition from corpora. In *Proceedings of the 18th Conference on Computational Linguistics*, pages 871–877.

Yokobayashi, H., Suganuma, A., and Taniguchi, R.-i. (2004). Generating candidates for rewriting based on an indicator of complex dependency and it' s application to a writing tool. *Journal of Information Processing Society of Japan*, 45(5):1451–1459 (In Japanese).

Yoshida, K., Ohno, T., Kato, Y., and Matsubara, S. (2014). Japanese word reordering integrated with dependency parsing. In *Proceedings of the 25th International Conference on Computational Linguistics*, pages 1186–1196.