# Multi-Perspective Document Revision

**Mana Ihori** and **Hiroshi Sato** and **Tomohiro Tanaka** and **Ryo Masumura**

NTT Computer and Data Science Laboratories, NTT Corporation

1-1 Hikarinooka, Yokosuka-Shi, Kanagawa 239-0847, Japan

`mana.ihori.kx@hco.ntt.co.jp`

## Abstract

This paper presents a novel multi-perspective document revision task. In conventional studies on document revision, tasks such as grammatical error correction, sentence reordering, and discourse relation classification have been performed individually; however, these tasks simultaneously should be revised to improve the readability and clarity of a whole document. Thus, our study defines multi-perspective document revision as a task that simultaneously revises multiple perspectives. To model the task, we design a novel Japanese multi-perspective document revision dataset that simultaneously handles seven perspectives to improve the readability and clarity of a document. Although a large amount of data that simultaneously handles multiple perspectives is needed to model multi-perspective document revision elaborately, it is difficult to prepare such a large amount of this data. Therefore, our study offers a multi-perspective document revision modeling method that can use a limited amount of matched data (i.e., data for the multi-perspective document revision task) and external partially-matched data (e.g., data for the grammatical error correction task). Experiments using our created dataset demonstrate the effectiveness of using multiple partially-matched datasets to model the multi-perspective document revision task.

## 1 Introduction

With the advance of natural language processing technology using deep learning, applications for writing support systems have been developed (Tsai et al., 2020; Ito et al., 2020). Such systems often implement a grammatical error correction task that corrects errors such as typos and mistakes in inflected verb forms (Rothe et al., 2021). It is easy for the reader to understand an error-free document, and the lack of errors can allow for smooth text communication. In addition, it is crucial to revise a document automatically because it is difficult to read one's writing objectively, and it is time-consuming for a third party to revise the document.

The document revision task has been studied in the natural language processing field by being broken down into partial tasks. Grammatical error correction tasks have been studied actively (Sawai et al., 2013; Mizumoto and Matsumoto, 2016; Junczys-Dowmunt and Grundkiewicz, 2016), and modeling with the sequence-to-sequence (seq2seq) model has achieved high performance with the advance of deep learning (Yuan and Briscoe, 2016; Junczys-Dowmunt et al., 2018; Rothe et al., 2021). In addition, tasks considering the relationship between sentences in a document are sentence ordering (Yin et al., 2019; Wang and Wan, 2019) and discourse relation classification (Liu et al., 2016; Dai and Huang, 2018). These tasks have achieved high performance using deep learning, as with the grammatical error correction task.

However, studies that simultaneously revise multiple perspectives have not been well considered. To advance writing support, it is important not only to correct grammatical errors in a single sentence but also to improve the readability and clarity of a whole document. For example, when we manually perform document revision, we attempt to correct grammatical errors, split a long sentence into shorter sentences, and consider the relationships between sentences, such as by reordering them to obtain a consistent order and by performing conjunction insertion. Accordingly, this paper addresses a novel multi-perspective document revision task that simultaneously considers various perspectives such as grammatical error correction, long sentence splitting, sentence reordering, and conjunction insertion, as shown in Figure 1.

To address the multi-perspective document revision task, we need to define this task and design a suitable dataset. In this paper, we define

6128

**Pre-revision document:**

Until now, the mainstream of **education** has focused on grammar and formal aspects, but this is somewhat counterproductive for elementary school students who **has** not yet formed the ability to think logically and that it is more effective for them to learn practical English. I believe that **studying English** in Japan should be practical. **Arithmetic is also crucial in elementary school.** **I should review English education in elementary schools.**

Sentence splitting    Word unification    Grammatical error correction    Sentence deletion    Sentence reordering

**Post-revision document:**

**I should review English education in elementary schools.** Until now, the mainstream of **English education** has focused on grammar and formal aspects. However, this is somewhat counterproductive for elementary school students who **have** not yet formed the ability to think logically. It is effective for them to learn practical English. **Therefore**, I believe that **English education** in Japan should be practical.
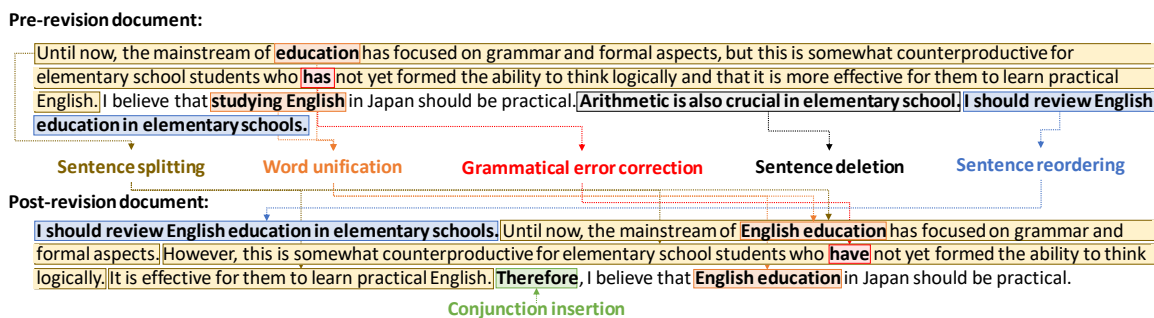
Conjunction insertion

Figure 1: Example of multi-perspective document revision task.

multi-perspective document revision as revising a document to improve its readability and clarity by considering multiple perspectives simultaneously. However, there is no dataset that simultaneously handles multiple perspectives because datasets in conventional studies handled only one perspective (Dahlmeier et al., 2013; Tanaka et al., 2020; Chen et al., 2016; Webber et al., 2019). Thus, we make a new Japanese multi-perspective document revision dataset that can simultaneously handle seven perspectives of readability and clarity of documents. To the best of our knowledge, this is the first attempt to construct a multi-perspective document revision dataset. This paper details how we constructed our dataset (see section 3).

To model the multi-perspective document revision task, it is necessary to consider long-range contexts of multiple sentences because some perspectives address the relationships between sentences. For example, sentence reordering and conjunction insertion tasks cannot be revised without considering the relationship between multiple sentences. Thus, we use the seq2seq model for modeling the multi-perspective document revision task by considering document-level information. Although our study uses the seq2seq model as with the grammatical error correction task, one difference is that we handle not a single sentence but a set of sentences.

The main difficulty in modeling the multi-perspective document revision task is simultaneously considering multiple perspectives. To address the multiple perspectives simultaneously and robustly, we should prepare a large amount of matched data (i.e., data for the multi-perspective document revision task). However, it is difficult to prepare such a large amount of data because manually writing and revising documents is time-consuming. Thus, we use a limited amount of matched data and a large amount of partially-matched data that handles individual partial tasks

to model a multi-perspective document revision task because this task is composed of multiple partial tasks. Preparing a large amount of partially-matched data is easy because some datasets exist, and others can be generated heuristically (e.g., for the conjunction insertion task, we can construct paired data by deleting and restoring conjunctions from existing documents). To effectively model the multi-perspective document revision task using both a matched dataset and multiple partially-matched datasets, we use seq2seq modeling with switching tokens (Ihori et al., 2021b). In our study, the switching tokens are used for distinguishing individual partial tasks in our multi-perspective document revision task. For example, when the grammatical error correction dataset is trained, we can use this dataset as the partially-matched dataset by switching the grammatical error correction task "on" and other tasks "off". Although the method using switching tokens is not new, applying the switching tokens for the task that can improve the performance of a matched task from multiple partially-matched tasks is new.

In our experiments using our created dataset, we use the grammatical error correction dataset and the conjunction insertion dataset as partially-matched datasets. Our results demonstrate that our modeling method can simultaneously revise multiple perspectives and effectively improve performance by using a matched dataset and multiple partially-matched datasets. Our main contributions are as follows:

- We define a novel multi-perspective document revision task that simultaneously considers multiple perspectives for writing support.

- We create a novel Japanese multi-perspective document revision dataset that can simultaneously handle seven perspectives and detail how we make it.

6129

- We present a multi-perspective document revision modeling method that takes advantage of the fact that this task is composed of multiple partial tasks and uses both a matched dataset and multiple partially-matched datasets.

## 2 Related Work

### 2.1 Modeling of partial task in document revision task

The partial tasks that compose a document revision task have been studied as individual tasks. First, grammatical error correction is the most typical task, and it corrects the errors in input text by deleting, inserting, and replacing words. For this task, studies have focused on sentence-level errors and performed error correction by using a seq2seq model to achieve high-performance (Yuan and Briscoe, 2016; Junczys-Dowmunt et al., 2018; Rothe et al., 2021). Also, synthetic training data generation is introduced to deal with paired-data scarcity in recently (Grundkiewicz et al., 2019; Kiyono et al., 2020; Rothe et al., 2021). However, it is difficult to generate synthetic data for the multi-perspective document revision task because it involves multiple partial tasks such as grammatical error correction, sentence reordering, and conjunction insertion. Next, there are the tasks that handle a set of sentences like the discourse relation classification task (Liu et al., 2016; Dai and Huang, 2018). This task predicts the relation class (e.g., contrast and causality) of two arguments and can help in writing coherent text by suggesting relationships between sentences. Our study adopts a conjunction insertion task similar to the discourse relation classification task but directly completes conjunctions in accordance with the relationship between sentences.

### 2.2 Modeling of multiple perspectives simultaneously

There are few studies to perform multiple perspectives using the seq2seq model. Lin et al. (2021) proposed document-level paraphrase generation task that simultaneously performs the sentence reordering and sentence paraphrasing tasks. In this conventional study, a pseudo dataset for a document-level paraphrase generation task was created, and the task was performed with a task-specific model architecture. Ihori et al. (2021b) proposed the method to perform disfluency deletion and punctuation restoration tasks simultaneously. To execute

| (1) | Correct the following mistakes. erroneous substitution, deletion, insertion, and kanji-conversion, syntactic errors, redundant expressions, style normalization, punctuation errors |
|-----|------|
| (2) | Split long sentences containing more than 60 characters. |
| (3) | Unify words with different expressions that have the same meaning. |
| (4) | If there is no subject, restore the subject by using words that have already been mentioned. |
| (5) | Change the sentence order if it is not appropriate. |
| (6) | Delete sentences that describe unrelated topics. |
| (7) | Insert correct conjunctions by considering the relationships between sentences. |

Table 1: Perspectives for Japanese multi-perspective document revision.

these two tasks simultaneously without preparing a matched dataset, switching tokens have been introduced into the seq2seq model. These conventional studies handle limited tasks in the document revision task, and this paper is the first study to consider more perspectives simultaneously than these studies.

## 3 Japanese Multi-perspective Document Revision Dataset

This section details a new dataset for a Japanese multi-perspective document revision task. The dataset contains paired data consisting of source and revised documents in Japanese. The source documents were written by Japanese crowd workers, and the reference documents were revised by two Japanese labelers. To revise documents, we defined seven perspectives that have been individually used in general document revision problems. Table 1 summarizes all perspectives, and Figure 2 shows an example of multi-perspective document revision that simultaneously uses several perspectives, correction of erroneous insertion and punctuation error, splitting of sentences, and conjunction insertion. To the best of our knowledge, this is the first dataset to address such multiple perspectives of the document revision task.

### 3.1 Perspectives

**(1) Error correction** This perspective includes the grammatical error correction task (Tanaka et al., 2020). Mistakes in a document need to be corrected because it is difficult to understand the document with errors. In this paper, we define eight Japanese-specific errors, erroneous substitution, deletion, insertion, and kanji-conversion, syntactic errors, re-

| Source document | ソーシャルメディアの発達**により**、私たちは好きな情報を簡単に得られるようになった。その一方で、日々膨大な情報にさらされることの弊害も出てきている。ソーシャルメディアでは、自身の趣味嗜好に**あ**沿った情報を得ようとすることが多い。これまでは新聞やテレビが情報源だったので全員が同じ情報に触れて対等に対話していた**が**、現在は無意識に自分に近い意見を全体の情報として捉えてしまう人もおり、情報が偏る。ソーシャルメディアは一見情報の宝庫に見えるが、視点を変えれば物事を都合よく**とらえる**為のツールなのかもしれない。 |
|---|---|
| Revised document | ソーシャルメディアの発達**で**、私たちは好きな情報を簡単に得られるようになった。その一方で、日々膨大な情報にさらされる弊害も出てきている。**また、**ソーシャルメディアでは、自身の趣味嗜好に沿った情報を得ようとすることが多い。これまでは新聞やテレビが情報源だったので、全員が同じ情報に触れて対等に対話していた。**しかし、**現在は、無意識に自分に近い意見を全体の情報として捉える人もおり、情報が偏る。**そのため、**ソーシャルメディアは一見情報の宝庫に見えるが、視点を変えれば、物事を都合よく**捉える**為のツールなのかもしれない。 |
| Translation | The development of social media has made it easier to get information. On the other hand, there can be difficulties in handling vast amounts of information. Also, in most cases, we only use social media to access our favorite types and sources of information. Previously, many people got the same information from newspapers and television, and thus, they could talk on an equal footing. Now, however, some people unknowingly treat their closely held opinions as complete information, so their information is biased. Therefore, social media seems to be a treasure trove of information, but it may also be a tool for maintaining biased information. |

Figure 2: Example of Japanese multi-perspective document revision dataset. The colors of characters correspond to perspectives in Figure 1.

dundant expressions, text style normalization, and punctuation errors. For example, in revising redundant expressions, we remove expressions with the same meaning or words that make sense without them (e.g., 一番最後→最後 # last, することができる→できる # can).

**(2) Sentence splitting**  Sentences more extended than 60 characters, which is the length of a typical sentence in Japanese, should be divided because long sentences decrease readability. We defined this by using specific numerical values to prevent different labelers from having other divisions.

**(3) Word unification**  To avoid confusion for the reader, words that have the same meaning in a document should be expressed in the same way. In addition, in Japanese documents, expressions "desu, masu" or "da, dearu," are used at the end of sentences, and these expressions need to be unified within the same document.

**(4) Subject restoration**  Throughout Japanese documents, the subject is often omitted; however, subject restoration may be necessary because a sentence without a subject may not correctly convey the intent of the writer to the reader. Thus, we restore the subject in sentences without a subject. Note that subject restoration should not be performed when the subject is clearly recognizable because consecutive occurrences of the same subject may reduce readability.

**(5) Sentence reordering**  This perspective includes the sentence ordering task (Barzilay and Lapata, 2008). A well-organized document with a logical structure is much easier for people to read and understand. Also, sentence order is important in constructing logical structures. Thus, we reorder sentences, in order for a document to have a logical order.

**(6) Sentences deletion**  A coherent document consists of multiple sentences describing a single topic, but a document that mixes multiple topics can inhibit the understanding of the reader. Thus, we delete sentences that describe distinctly different topics from a document.

**(7) Conjunction insertion**  This perspective includes the discourse relation classification task (Liu et al., 2016). Discourse relations support a set of sentences to form a coherent document. Also, the conjunction has a role in showing these relations and serves as a natural means to connect sentences. Thus, we restore the conjunction in documents to improve readability. We created a list of conjunctions that show their kinds and roles, and we asked labelers to select from this list.

### 3.2 Dataset specifications

**Source documents:**  To make the source documents, we hired 161 Japanese workers through a crowdsourcing service and asked them to write paragraph-level documents in Japanese. The documents had an essay-style structure because Japanese schools teach how to write essays; thus, we expected that many workers could write essays at the same level. First, we showed the workers 48 themes, and they each selected 1-15 themes. The 48 themes were chosen by the crowdsourcing company from actual themes that were used for exam essays in Japan. Next, the workers wrote paragraph-level documents, each of which contained 200-300 characters and four or more sentences. We could

|  |  | # of documents | # of sentences |
|---|---|---|---|
| Training | Input | 5,000 | 26,477 |
|  | Output | 5,000 | 28,158 |
| Validation | Input | 554 | 2,922 |
|  | Output | 554 | 3,128 |
| Test | Input | 1,121 | 6,054 |
|  | Output | 2,242 | 12, 831 |

Table 2: Details of multi-perspective document revision dataset.

revise these source documents with multiple perspectives including the relationship between sentences by using this source document because they consisted of multiple sentences and had a coherent topic. Each worker wrote 1-15 documents, and the time limit for writing each one was 15 minutes. Although we asked workers to be careful about typos, we also asked them not to strive for perfection.

**Revised documents:** To revise the source documents, we hired two Japanese labelers. One labeler had a license as a Japanese language teacher, while the other labeler received guidance on revising the document. For the multi-perspective document revision task, we should simultaneously handle multiple perspectives to improve the readability and clarity of a document. Thus, we asked them to follow the revision guidelines listed in Table 1 to ensure that they could consider revising from multiple perspectives. We expected that the labelers would be able to revise documents with equivalent quality by following the guidelines. Note that they did not necessarily have to consider all perspectives simultaneously but were only to make these revisions if there were any mistakes or unnatural points. The collected data was divided into a training set, a validation set, and a test set. Table 2 details the resulting dataset for the Japanese multi-perspective document revision task.

### 3.3 Analysis

We investigate how much of the source documents were revised. To investigate the revision, we employ and measure Levenshtein distance (Levenshtein et al., 1966), which can measure the edit distance between the source and revised documents. Figure 3 shows the distribution of Levenshtein distance for all paired data in our created multi-perspective document revision dataset. In Figure 3, the Levenshtein distance of 17 was the most common, and the paired data with the distance were revised by correcting punctuation errors and inserting one or two conjunctions. In addition, as the Leven-
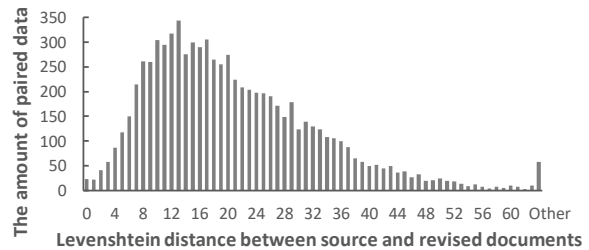


Figure 3: Distribution of Levenshtein distance for paired data in multi-perspective document revision dataset.

shtein distance increased, many perspectives were corrected simultaneously, such as reordering sentences, restoring subjects, and splitting sentences. However, the distribution of the various perspectives is unbalanced because this dataset contains more conjunction insertion and error correction than other perspectives.

## 4 Multi-perspective Document Revision Models

### 4.1 Strategy

To build a multi-perspective document revision model, we use the matched dataset created in section 3 and multiple partially-matched datasets that handle the partial tasks of the multi-perspective document revision task. Our strategy in using these datasets jointly is to incorporate multiple "on-off" switches into the seq2seq model. These switches can be implemented by using switching tokens (Ihori et al., 2021b). A switching token represents the "on" state (the target task) or "off" state (not the target task) for each task. In addition, a model that introduces switching tokens can explicitly distinguish the multi-perspective document revision task and each partial task.

Figure 4 shows an example of training and decoding the multi-perspective document revision model using switching tokens. In the figure, we use three datasets, a matched dataset, a grammatical error correction (gec) dataset, and a conjunction insertion (ci) dataset, to train the multi-perspective document revision model. In addition, we use three switching tokens, $s_1 \in \{[\text{gec\_on}], [\text{gec\_off}]\}$, $s_2 \in \{[\text{ci\_on}], [\text{ci\_off}]\}$, and $s_3 \in \{[\text{other\_on}], [\text{other\_off}]\}$. We specify the "other" task because the multi-perspective document revision task handles other perspectives in addition to the grammatical error correction and conjunction insertion tasks, as listed in Table 1. These switching tokens are used as inputs of the decoder network in given
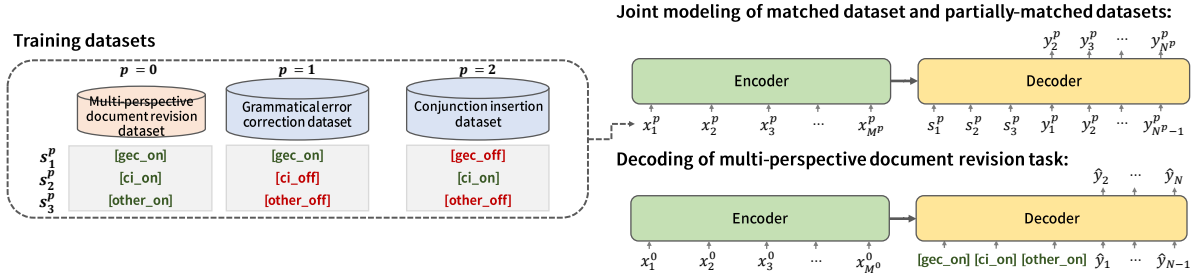
Figure 4: Example of training and decoding multi-perspective document revision model using switching tokens.

contexts. In the training phase, all three datasets are trained jointly by distinguishing each task with three switching tokens. In the decoding phase, the model performs the multi-perspective document revision task by feeding switching tokens, [gec_on], [ci_on], and [other_on]. Note that we can also perform the grammatical error correction or conjunction insertion tasks by feeding appropriate switching tokens.

## 4.2 Modeling method

We define a source document as $\boldsymbol{X} = \{x_1, \cdots, x_m, \cdots, x_M\}$ and a revised document as $\boldsymbol{Y} = \{y_1, \cdots, y_n, \cdots, y_N\}$, where $M$ and $N$ are the amount of tokens in source and revised documents, respectively. $x_m$ and $y_n$ are tokens that include not only characters or words but also punctuation marks. In this paper, we handle a set of sentences, so $\boldsymbol{X}$ and $\boldsymbol{Y}$ have multiple sentences. Thus, we introduce [CLS] token at the beginning of sentences into all datasets to distinguish each sentence in documents.

The multi-perspective document revision model predicts the generation probabilities of a revised document $\boldsymbol{Y}$ given a source document $\boldsymbol{X}$ and switching tokens $s_{1:T} = \{s_1, \cdots, s_t, \cdots, s_T\}$, where $T$ is the total number of partial tasks that are included in each partially-matched dataset. The generation probability of $\boldsymbol{Y}$ is defined as

$$P(\boldsymbol{Y}|\boldsymbol{X}, s_{1:T}; \boldsymbol{\Theta}) \qquad (1)$$
$$= \prod_{n=1}^{N} P(y_n|y_{1:n-1}, \boldsymbol{X}, s_{1:T}; \boldsymbol{\Theta}),$$

where $y_{1:n-1} = \{y_1, \cdots, y_{n-1}\}$, and $\boldsymbol{\Theta}$ represents the trainable parameters. $s_t$ is the $t$-th switching token represented as

$$s_t \in \{[t-\text{th task\_on}], [t-\text{th task\_off}]\}. \quad (2)$$

In this paper, we use Transformer pointer-generator networks (Deaton, 2019) for this model-

ing. These networks are effective for monolingual translation tasks because they contain a copy mechanism that copies tokens from a source text to help generate infrequent tokens. Note that our method does not change the architecture of a transformer pointer-generator network but merely adds switching tokens to the model input.

**Pre-training:** In this paper, we use a MAsked Pointer-Generator Network (MAPGN) (Ihori et al., 2021a) as self-supervised pre-training for the seq2seq model because it is a suitable pre-training method for pointer-generator networks. In MAPGN, the pointer-generator network is pre-trained by predicting a sentence fragment $y_{a:b}$ given a masked sequence $\boldsymbol{Y}_{/a:b}$. Here, $\boldsymbol{Y}_{/a:b}$ denotes a fragment in which positions a to b are masked, and $y_{a:b}$ denotes a sentence fragment of $\boldsymbol{Y}$ from a to b. The model parameter set can be optimized from unpaired dataset $\mathcal{D}_\text{u}$ that is consisted of a set of sentences. The training loss function $\mathcal{L}$ is defined as

$$\mathcal{L} = - \sum_{\boldsymbol{Y} \in \mathcal{D}_\text{u}} \log P(y_{a:b}|y_{a-1}, \boldsymbol{Y}_{/a:b}; \boldsymbol{\Theta}), \qquad (3)$$
$$= - \sum_{\boldsymbol{Y} \in \mathcal{D}_\text{u}} \sum_{t=a}^{b} \log P(y_t|y_{a-1:t-1}, \boldsymbol{Y}_{/a:b}; \boldsymbol{\Theta}).$$

Note that all switching tokens have to be included in the vocabulary during pre-training.

**Fine-tuning:** During fine-tuning, the matched dataset $\mathcal{D}_\text{m}$, and multiple partially-matched datasets $\{\mathcal{D}_\text{pm}^1, \cdots, \mathcal{D}_\text{pm}^p, \cdots, \mathcal{D}_\text{pm}^P\}$ are trained jointly in a single model. $P$ is the number of partially-matched datasets. The training loss function $\mathcal{L}$ is defined as

$$\mathcal{L} = \mathcal{L}_\text{m} + \sum_{p=1}^{P} \mathcal{L}_\text{pm}^p, \qquad (4)$$

where $\mathcal{L}_{\mathtt{m}}$ is the loss function against the main task and it is computed from

$$\mathcal{L}_{\mathtt{m}} = -\sum_{(\boldsymbol{X}^0, \boldsymbol{Y}^0) \in \mathcal{D}_{\mathtt{m}}} \log P(\boldsymbol{Y}^0 | \boldsymbol{X}^0, s_{1:T}^0; \boldsymbol{\Theta}), \tag{5}$$

where $s_{1:T}^0 = \{s_1^0, \cdots, s_T^0\}$ are switching tokens and $s_t^0$ is represented as

$$s_t^0 = [t-\text{th } \mathtt{task\_on}]. \tag{6}$$

$\mathcal{L}_{\mathtt{pm}}^p$ is the loss function against the $p$-th partially-matched dataset and it is computed from

$$\mathcal{L}_{\mathtt{pm}}^p = -\sum_{(\boldsymbol{X}^p, \boldsymbol{Y}^p) \in \mathcal{D}_{\mathtt{pm}}^p} \log P(\boldsymbol{Y}^p | \boldsymbol{X}^p, s_{1:T}^p; \boldsymbol{\Theta}), \tag{7}$$

where $s_{1:T}^p = \{s_1^p, \cdots, s_T^p\}$ are switching tokens and $s_t^p$ is represented as

$$s_t^p = \begin{cases} [t-\text{th } \mathtt{task\_on}] & \text{if } t\text{-th task in } \mathcal{D}_{\mathtt{pm}}^p, \\ [t-\text{th } \mathtt{task\_off}] & \text{otherwise.} \end{cases} \tag{8}$$

**Decoding:** The decoding problem using switching tokens is defined as

$$\hat{\boldsymbol{Y}} = \arg\max_{\boldsymbol{Y}} P(\boldsymbol{Y} | \boldsymbol{X}, s_{1:T}; \boldsymbol{\Theta}). \tag{9}$$

The model can perform the multi-perspective document revision task or each partial task in accordance with the given switching tokens.

## 5 Experiments

We experimentally evaluated the effectiveness of this modeling method that can use both a matched dataset and multiple partially-matched datasets.

### 5.1 Dataset

In this paper, we use two partial tasks, the grammatical error correction (gec) and conjunction insertion (ci) tasks, to build a multi-perspective document revision model. Accordingly, we use three datasets, a multi-perspective document revision dataset described in section 3, a Japanese grammatical error correction dataset (Tanaka et al., 2020), and a conjunction insertion dataset. We made the conjunction insertion dataset by deleting and restoring conjunctions from the Japanese Wiki-40b dataset (Guo et al., 2020), which is a high-quality processed Wikipedia dataset. To make this dataset

|  |  | # of documents | # of sentences |
|---|---|---|---|
| Training | a | 5,000 | 26,477 |
|  | b | - | 506,786 |
|  | c | 90,000 | 533,422 |
| Validation | a | 554 | 2,922 |
|  | b | - | 8,542 |
|  | c | 10,000 | 59,396 |
| Test | a | 1,121 | 6,054 |
|  | b | - | 8,542 |
|  | c | 1,000 | 6,026 |
| switch | a | [gec_on][ci_on][other_on] | |
|  | b | [gec_on][ci_off][other_off] | |
|  | c | [gec_off][ci_on][other_off] | |

a. Multi-perspective document revision dataset
b. Japanese grammatical error correction dataset
c. Conjunction insertion dataset

Table 3: Details of a matched dataset and two partially-matched datasets.

from paragraph-level documents, we divided the dataset into paragraphs and extracted the documents that contained conjunctions. In addition, we use unpaired 880k paragraph-level documents for self-supervised pre-training, and these documents, which were prepared from the Wiki-40B dataset, were not used in the conjunction insertion dataset. The details of these datasets are listed in Table 3, where "switch" refers to switching tokens. For training and decoding, we use three switching tokens for the gec task, the ci task, and other tasks. The multi-perspective document revision model can also perform the gec and ci tasks by feeding appropriate switching tokens, so we evaluate the performance of each partial task using each test set. Note that the Japanese grammatical error correction dataset does not have documents because this dataset consists of single sentences.

### 5.2 Setup

For evaluation purposes, we constructed 11 Transformer-based pointer-generator networks. We use the three datasets combined in different ways, each dataset only, a matched dataset and each partially-matched dataset, and the three datasets jointly. Then, we construct the models with and without switching tokens. In addition, we apply the pre-training to the models that use a matched dataset only and the three datasets jointly with switching tokens.

We used the following configurations. The encoder and decoder had a 4-layer and 2-layer transformer block with 512 units. The output unit size (corresponding to the number of tokens in the pre-training data) was set to 12,773. To train the

6134

|         |     | Dataset   | GLEU  | $F_{0.5}$ | C-F1  |
|---------|-----|-----------|-------|-----------|-------|
| Source  |     | -         | 0.886 | 0         | 0     |
| Baseline| (1) | a         | 0.857 | 0.198     | 0.193 |
|         | (2) | + PT      | 0.884 | 0.321     | 0.211 |
| + Datasets | (3) | a + b  | 0.863 | 0.189     | 0.164 |
|         | (4) | a + c     | 0.881 | 0.155     | 0.101 |
|         | (5) | a + b + c | 0.883 | 0.236     | 0.205 |
| + Switch | (6) | a + b    | 0.887 | 0.278     | 0.163 |
|         | (7) | a + c     | 0.888 | 0.234     | 0.214 |
|         | (8) | a + b + c | 0.889 | 0.282     | 0.270 |
|         | (9) | + PT      | **0.892** | **0.333** | **0.274** |

Table 4: Results of multi-perspective document revision task. "Source" row indicates results for source documents in multi-perspective document revision task dataset.

| Task | Dataset   | GLEU      | $F_{0.5}$ | C-F1  |
|------|-----------|-----------|-----------|-------|
| gec  | b         | **0.943** | **0.635** | -     |
|      | a + b + c | 0.932     | 0.613     | -     |
|      | + switch  | **0.943** | 0.630     | -     |
| ci   | c         | 0.964     | 0.198     | 0.230 |
|      | a + b + c | 0.966     | 0.207     | 0.222 |
|      | + switch  | **0.967** | **0.239** | **0.263** |

Table 5: Results of grammatical error correction (gec) and conjunction insertion (ci) tasks.

Transformer pointer-generator networks, we used the RAdam optimizer (Liu et al., 2019) and label smoothing (Lukasik et al., 2020) with a smoothing parameter of 0.1. We set the mini-batch size to 32 documents and the dropout rate in each Transformer block to 0.1. All trainable parameters were initialized randomly, and we used characters as tokens. In the pre-training, we set the number of masked tokens to roughly 50% of an input sequence. For decoding, we used the beam search algorithm with a beam size of 4.

## 5.3 Results

Tables 4 and 5 show the results for the 11 Transformer pointer-generator networks. In these tables, a, b, and c represent the multi-perspective document revision dataset, the Japanese grammatical error correction dataset, and the conjunction insertion dataset. Also, "switch" and "PT" indicate switching tokens and pre-training, respectively. We used automatic evaluation scores in terms of two metrics: GLEU (Napoles et al., 2015) and $F_{0.5}$. Specifically, we calculated these metrics for characters in generated documents and used 4-grams for GLEU. In addition, we also calculated the F1 score for conjunction insertion, denoted as C-F1, to evaluate the performance of the conjunction insertion task. We evaluated whether the system could insert conjunctions with the correct meaning be-

cause multiple conjunctions have the same sense (e.g., "but" and "however").

**Results of multi-perspective document revision task:** In Table 4, when the number of partially-matched datasets increased, the performance improved. This indicated that switching-token-based joint modeling that trains a matched dataset and multiple partially-matched datasets using switching tokens is effective for modeling the multi-perspective document revision task. Therefore, it is important for switching-token-based joint modeling to distinguish tasks because the scores with switching tokens were higher than those without switching tokens. In addition, when we compared the results of lines (8) with (9), the results with pre-training outperformed those without pre-training. This indicates that switching-token-based joint modeling can be effectively applied after performing self-supervised pre-training.

Figure 5 shows that the generation examples of lines (2), (5), and (8) in Table 4. In the figure, the generation example in (2) shows that conversion errors were decreased, but task-specific generation, like a conjunction insertion, was not performed well. On the other hand, the generation example in (8) shows that task-specific generation performed better than (2). Thus, we suppose it is difficult to improve the performance of the multi-perspective document revision task by applying pre-training alone. Here, the generation example in (5) shows it has more errors than (8) in a task-specific generation. These facts indicate that the switching tokens are important for joint training of the matched and multiple partially-matched tasks.

**Results of partial tasks:** Table 5 shows that in grammatical error correction, the performance of individual modeling and switching-token-based joint modeling were not significantly different. However, the performance of joint modeling without switching tokens under-performed that of the individual modeling. For the conjunction insertion task, the results of joint modeling outperformed those of individual modeling. Also, the results of joint modeling with switching tokens outperformed those without switching tokens. Therefore, these results indicated that switching-token-based joint modeling could improve the performance of the multi-perspective document revision task without impairing the performance of each task. Note that this study aims to improve the performance

| | |
|---|---|
| Input | i. …相手理解する (I understand you)，…つながらない (disconnect)，…成功をおさめる (achieve success) <br> ii. …現代社会において、…「共依存」の関係が問題となっている。古代において敵が多い環境では、自分の命は自分で守る… 大人になった証でもあった。誰かに依存することは多くの人の命を脅かす危険があったのである。 <br> (... In today's society, ... "co-dependent relationships" have become a problem. In ancient times and in an environment full of enemies, one's life was one's own ... a sign of adulthood. To be dependent on someone else was to risk the lives of many.) |
| Reference | i. …相手を理解する，…繋がらない，…成功を収める <br> ii. …現代社会において、…「共依存」の関係が問題となっている。古代において敵が多い環境では、自分の命は自分で守る… 大人になった証でもあった。つまり、誰かに依存することは多くの人の命を脅かす危険があったのである。 <br> (... In today's society, ... "co-dependent relationships" have become a problem. In ancient times and in an environment full of enemies, one's life was one's own ... a sign of adulthood. In other words, to be dependent on someone else was to risk the lives of many.) |
| (2) | i. …相手理解する，…つながらない，…成功をおさめる <br> ii. …現代社会で、…「共依存」の関係が問題となっている。古代において、敵が多い環境では、自分の命は自分で守る…大人になった証でもあった。そのため、誰かに依存することは、多くの人の命を脅かす危険があったのである。 <br> (... In today's society, ... "co-dependent relationships" have become a problem. In ancient times and in an environment full of enemies, one's life was one's own ... a sign of adulthood. Thus, to be dependent on someone else was to risk the lives of many.) |
| (5) | i. …相手理解する，…繋がらない，…成功をおさめる <br> ii. …現代社会で、…「共依存」の関係が問題となっている。なぜなら、古代において敵が多い環境では、自分の命は自分で守る…大人になった証でもあったからだ。したがって、誰かに依存することは多くの人の命を脅かす危険があったのである。 <br> (... In today's society, ... "co-dependent relationships" have become a problem. Since in ancient times and in an environment full of enemies, one's life was one's own ... a sign of adulthood. Therefore, to be dependent on someone else was to risk the lives of many.) |
| (8) | i. …相手を理解する，…繋がらない，…成功を収める <br> ii. …現代社会で、…「共依存」の関係が問題となっている。古代において敵が多い環境では、自分の命は自分で守る…大人になった証でもあった。つまり、誰かに依存することは多くの人の命を脅かす危険があったのである。 <br> (... In today's society, ... "co-dependent relationships" have become a problem. In ancient times and in an environment full of enemies, one's life was one's own ... a sign of adulthood. In other woeds, to be dependent on someone else was to risk the lives of many.) |

i: Correction focused on grammatical error correction, ii: Correction focused on conjunction insertion

Figure 5: Generation examples of lines (2), (5), and (8) in Table 4.

of the matched task by simultaneously training the matched task and multiple partially-matched tasks, so we do not aim to improve the performance of individual partially-matched tasks.

**Conflict results in Tables 4 and 5:** The results of lines (4), (5), (7), and (8) in Table 4 show irrelevant dataset b brings more prominent improvement on C-F1. However, In Table 5, after introducing dataset b in ci task, C-F1 shows an obvious downward trend. We suppose these results are dependent on the target tasks. First, we focus on the results of the multi-perspective document revision task in Table 4. Since this task requires multiple tasks to map simultaneously from a source document to a revised document, this task is more complex than handling a single task and requires a large amount of training data. We think the results of each task in the multi-perspective document revision task are improved by simultaneously training the dataset for different tasks included in this task. This reason is that the performance of the seq2seq mapping is improved due to increasing the amount of training data that partially handles the multi-perspective document revision task. Next, we focus on the results of a single task in Table 5. Here, the ci and gec tasks are related tasks for the multi-perspective document revision task, but each task is unrelated. Thus, in the ci task, adding datasets for unrelated tasks without switching tokens may have degraded performance for this task because data from unrelated tasks may have become noise.

## 6 Conclusion

In this paper, we proposed a novel multi-perspective document revision task that revises multiple perspectives simultaneously to improve the readability and clarity of a document. We created a dataset that simultaneously addresses seven perspectives by using a crowdsourcing service. In addition, to model the multi-perspective document revision task, we presented a seq2seq modeling method with multiple "on-off" switches. This method allowed us to effectively use a multi-perspective document revision dataset and partially-matched datasets, the grammatical error correction and the conjunction insertion datasets. The experimental results obtained using our created dataset demonstrated that using switches is important for modeling the multi-perspective document revision task. In our future work, we will increase the number of partial tasks (e.g., sentence reordering) and develop a model architecture that is suitable for handling much longer documents.

## References

Regina Barzilay and Mirella Lapata. 2008. Modeling local coherence: An entity-based approach. *Computational Linguistics*, pages 1–34.

Xinchi Chen, Xipeng Qiu, and Xuanjing Huang. 2016. Neural sentence ordering. *arXiv preprint arXiv:1607.06952*.

Daniel Dahlmeier, Hwee Tou Ng, and Siew Mei Wu. 2013. Building a large annotated corpus of learner

english: The nus corpus of learner english. In *Proceedings of the eighth workshop on innovative use of NLP for building educational applications*, pages 22–31.

Zeyu Dai and Ruihong Huang. 2018. Improving implicit discourse relation classification by modeling inter-dependencies of discourse units in a paragraph. In *Proc. Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, pages 141–151.

John Deaton. 2019. Transformers and pointer-generator networks for abstractive summarization.

Roman Grundkiewicz, Marcin Junczys-Dowmuntz, and Kenneth Heafield. 2019. Neural grammatical error correction systems with unsupervised pre-training on synthetic data. In *Proc. Workshop on Innovative Use of NLP for Building Educational Applications (BEA Workshop)*, pages 252–263.

Mandy Guo, Zihang Dai, Denny Vrandečić, and Rami Al-Rfou. 2020. Wiki-40b: Multilingual language model dataset. In *Proc. Language Resources and Evaluation Conference (LREC)*, pages 2440–2452.

Mana Ihori, Naoki Makishima, Tomohiro Tanaka, Akihiko Takashima, Shota Orihashi, and Ryo Masumura. 2021a. Mapgn: Masked pointer-generator network for sequence-to-sequence pre-training. In *Proc. International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 7563–7567.

Mana Ihori, Naoki Makishima, Tomohiro Tanaka, Akihiko Takashima, Shota Orihashi, and Ryo Masumura. 2021b. Zero-Shot Joint Modeling of Multiple Spoken-Text-Style Conversion Tasks Using Switching Tokens. In *Proc. International Speech Communication Association (Interspeech)*, pages 776–780.

Takumi Ito, Tatsuki Kuribayashi, Masatoshi Hidaka, Jun Suzuki, and Kentaro Inui. 2020. Langsmith: An interactive academic text revision system. In *Proc. Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 216–226.

Marcin Junczys-Dowmunt and Roman Grundkiewicz. 2016. Phrase-based machine translation is state-of-the-art for automatic grammatical error correction. In *Proc. Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1546–1556.

Marcin Junczys-Dowmunt, Roman Grundkiewicz, Shubha Guha, and Kenneth Heafield. 2018. Approaching neural grammatical error correction as a low-resource machine translation task. In *Proc. Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, pages 595–606.

Shun Kiyono, Jun Suzuki, Masato Mita, Tomoya Mizumoto, and Kentaro Inui. 2020. An empirical study of incorporating pseudo data into grammatical error correction. In *Proc. Conference on Empirical Methods in Natural Language Processing and International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 1236–1242.

Vladimir I Levenshtein et al. 1966. Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet physics doklady*, pages 707–710.

Zhe Lin, Yitao Cai, and Xiaojun Wan. 2021. Towards document-level paraphrase generation with sentence rewriting and reordering. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 1033–1044.

Liyuan Liu, Haoming Jiang, Pengcheng He, Weizhu Chen, Xiaodong Liu, Jianfeng Gao, and Jiawei Han. 2019. On the variance of the adaptive learning rate and beyond. In *Proc. International Conference on Learning Representations (ICLR)*.

Yang Liu, Sujian Li, Xiaodong Zhang, and Zhifang Sui. 2016. Implicit discourse relation classification via multi-task neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

Michal Lukasik, Srinadh Bhojanapalli, Aditya Menon, and Sanjiv Kumar. 2020. Does label smoothing mitigate label noise? In *Proc. the International Conference on Machine Learning (ICML)*, pages 6448–6458.

Tomoya Mizumoto and Yuji Matsumoto. 2016. Discriminative reranking for grammatical error correction with statistical machine translation. In *Proc. Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, pages 1133–1138.

Courtney Napoles, Keisuke Sakaguchi, Matt Post, and Joel Tetreault. 2015. Ground truth for grammatical error correction metrics. In *Proc. Annual Meeting on Association for Computational Linguistics (ACL)*, pages 588–593.

Sascha Rothe, Sebastian Krause, Jonathan Mallinson, Eric Malmi, and Aliaksei Severyn. 2021. A simple recipe for multilingual grammatical error correction. In *Proc. Annual Meeting of the Association for Computational Linguistics and the International Joint Conference on Natural Language Processing (ACL-IJCNLP)*, pages 702–707.

Yu Sawai, Mamoru Komachi, and Yuji Matsumoto. 2013. A learner corpus-based approach to verb suggestion for ESL. In *Proc. Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 708–713.

Yu Tanaka, Yugo Murawaki, Daisuke Kawahara, and Sadao Kurohashi. 2020. Building a Japanese typo dataset from Wikipedia's revision history. In *Proc. Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 230–236.

Chung-Ting Tsai, Jhih-Jie Chen, Ching-Yu Yang, and Jason S. Chang. 2020. LinggleWrite: a coaching system for essay writing. In *Proc. Annual Meeting of*

6137

*the Association for Computational Linguistics (ACL)*, pages 127–133.

Tianming Wang and Xiaojun Wan. 2019. Hierarchical attention networks for sentence ordering. In *Proc. AAAI Conference on Artificial Intelligence (AAAI)*, pages 7184–7191.

Bonnie Webber, Rashmi Prasad, Alan Lee, and Aravind Joshi. 2019. The penn discourse treebank 3.0 annotation manual. *Philadelphia, University of Pennsylvania*.

Yongjing Yin, Linfeng Song, Jinsong Su, Jiali Zeng, Chulun Zhou, and Jiebo Luo. 2019. Graph-based neural sentence ordering. *arXiv preprint arXiv:1912.07225*.

Zheng Yuan and Ted Briscoe. 2016. Grammatical error correction using neural machine translation. In *Proc. Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, pages 380–386.