Building a Corpus of Manually Revised Texts from Discourse Perspective

Ryu Iida, Takenobu Tokunaga

Graduate School of Information Science and Engineering, Tokyo Institute of Technology
2-12-1 Ōokayama, Meguro-ku, Tokyo 152-8552, Japan
{ryu-i,take}@cl.cs.titech.ac.jp

Abstract

This paper presents building a corpus of manually revised texts which includes both before and after-revision information. In order to create such a corpus, we propose a procedure for revising a text from a discourse perspective, consisting of dividing a text to discourse units, organizing and reordering groups of discourse units and finally modifying referring and connective expressions, each of which imposes limits on freedom of revision. Following the procedure, six revisers who have enough experience in either teaching Japanese or scoring Japanese essays revised 120 Japanese essays written by Japanese native speakers. Comparing the original and revised texts, we found some specific manual revisions frequently occurred between the original and revised texts, e.g. ‘thesis’ statements were frequently placed at the beginning of a text. We also evaluate text coherence using the original and revised texts on the task of pairwise information ordering, identifying a more coherent text. The experimental results using two text coherence models demonstrated that the two models did not outperform the random baseline.

Keywords: essay revision, text coherence, discourse

1. Introduction

The research of NLP applications for improving student’s writing skills has grown rapidly in recent years (Dale and Kilgarriff, 2010), as one of education-oriented applications of NLP. The existing studies on these applications have mainly focused on detecting and correcting grammatical and spelling errors (Brockett et al., 2006; Hermet and Désilets, 2009; Leacock et al., 2010; Park and Levy, 2011). On the other hand, there has been growing need for applications taking into account discourse coherence of a text, e.g. automatic essay scoring (Shermis and Burstein, 2003) and essay revision. They are important because of the difficulty of consistent essay scoring by human evaluators. Furthermore, such applications are sometimes required to provide comprehensible reasons of evaluation scores and revision of the text for language learners. For example, learners of Japanese language sometimes use less cohesive referring expressions, e.g. excessive use of ellipses. The comprehensible explanation for appropriate usage of (zero) anaphors is required for the learners. However, generating such explanation has been less studied because the existing systems of writing support and text revision have less focused on the treatment of discourse processing.

Against this background, we aim at developing a technique of automatically revising a text from discourse perspective, taking Japanese as a target language. To this end, the following two approaches, which are complementary to each other, can be considered.

1. A rational approach develops a theory of realising coherent texts from theoretical perspectives of discourse.

2. An empirical approach creates manually revised texts for quantitative analysis and then develops a text revision model based on clues acquired from the analysis.

The former approach focuses on integration of various clues proposed in the past discourse studies (e.g. Centering Theory (Grosz et al., 1995) and Rhetorical Structure Theory (RST) (Mann and Thompson, 1988)) into the target problem (e.g. essay scoring or essay revision), relying on the notions of discourse theories (e.g. centers introduced in Centering Theory and discourse relations defined in RST). However, the effectiveness of those theoretically motivated clues on revising texts has not been studied comprehensively.

In this work, we take the latter empirical approach to tackle the task of text revision for achieving better discourse coherence in the revised texts. In order to empirically investigate the discourse characteristics during text revision, we first build a corpus including both before- and after-revision texts. Without explicit revision guidelines, human revisers tend to correct only superficial grammatical and spelling errors to make texts at least legitimate. More drastic modification involving rearrangement of sentence order and using alternate referring expressions for improving coherence remains to be performed. This would be because the revisers have too diverse possible modification operations to choose one, and they try to avoid accidental change of text contents against the author's original intention. For this reason, we design an explicit procedure for prompting a human reviser to take into account discourse coherence during her revision.

2. A procedure of revising texts

As described in Section 1, revisers tend to prioritise grammatical and spelling errors correction when revising texts. This tendency has an advantage that the revision does not change the content of a text originally intended by the author. However, there might still remain room for improvements in text coherence. For instance, changing the order of sentences, introducing appropriate conjunctives and furthermore supplementing additional sentences in the original text would contribute to improving text coherence. These operations should be also encouraged to adopt during text revision. Because introducing new sentences in-
volves difficult content selection issues, which depend on the background knowledge of a reviser, we exclude this option in this work. We eventually adopt the following three steps for revising texts, aiming at replicating these steps based on NLP techniques, especially automatic discourse analysis.

1) Dividing a text to discourse units: First, as a preprocessing of manual revision, a text is automatically decomposed into a series of discourse units (e.g. clauses), which roughly represents a logical proposition. In addition, discourse entities topicalised with case marker wa (topic) are off-topicalised by replacing with either go (subj), o (obj) or ni (obj) to neutralise its information status (e.g. old/new information). The conjugate form of the end phrase in each discourse unit is normalised. Due to the lack of reviser’s knowledge relating to natural language processing, we expected that automatically dividing a text to discourse units is rather consistent compared to manual discourse unit segmentation by revisers. For this reason, we employed a model of discourse unit segmentation based on the annotated data instead of manually segmenting discourse units.

2) Organising and reordering statement units: A reviser makes a cluster of adjacent discourse units representing a discourse function in the text (e.g. “thesis” and “elaboration”). We call this cluster the statement unit. The statement units are sequentially numbered from the beginning of the original text, e.g. SU0, SU1 .... The reviser then moves each statement unit to a more appropriate position to make the revised text more coherent. Furthermore, annotating a function label to each statement unit makes the reviser consider the discourse function of each statement unit within the text. The definition of the function labels for statement units in shown in Table 1, which is an extension of the work by Persing et al. (2010). In addition, tightly related statement units are assigned coindexed function labels, as the statement units 1 and 3 in Table 2. They share index 1 in their function label main, and elab.

3) Modifying referring and connective expressions: Reordering statement units might cause an incoherent text due to inappropriate referring and connective expressions. Therefore, the reviser modifies referring and connective expressions if needed for making the text more coherence after reordering statement units. To keep track of these modifications, we employed the notation of `(original expression/revised expression)`.

For each step, the reviser can refer to the original text if needed so as not to change the author’s original intention.

### Table 1: Statement unit labels

| label | name     | description                                                                 |
|-------|----------|------------------------------------------------------------------------------|
| pro   | prompt   | restates the prompt given to the author and contains no new material or opinions |
| tran  | transition | shifts the focus to new topics but contains no meaningful information         |
| the   | thesis   | states the author’s position on the topic for which she is arguing            |
| main  | main idea | asserts reasons and foundation arguments that support the thesis               |
| elab  | elaboration | further explains reasons and ideas but contains no evidence or examples       |
| supp  | support  | provides evidence and examples to support the claims made in other statements |
| conc  | conclusion | summarises and concludes the entire argument or one of the main ideas         |
| reb   | rebuttal | considers counter-arguments that contrast with the thesis or main ideas       |
| solu  | solution | puts to rest the questions and problems brought up by counter-arguments       |
| sugg  | suggestion | proposes solutions to the problems brought up by the argument                 |
| back  | background | states the background of main ideas                                           |

The selection of texts for manual revision is important on the task of automatic text revision because this work focuses on the quality of text organisation and factors relevant to the quality. We used 120 Japanese essays written by Japanese native speakers collected by Usami (2009). The essays were written by high school students on prompt “state your opinion in 800 characters about introducing English education into Japanese elementary schools.” Since the essays need to be informative and persuasive for readers on the controversial topic, the organisation of text is particularly important. Thus, we consider them a good material for our research purpose.

For preprocessing (step 1) in the procedure described in Section 2, we create a maximum entropy classifier which judges whether a bunsetsu-unit in a sentence is the end of a discourse unit. This problem setting is reasonable in Japanese because the right most bunsetsu-unit in a discourse unit, typically consisting of a predicate and its modality particles, gives a strong clue for the discourse unit boundary. We annotated a sub-corpus of a Japanese balanced corpus, BCCWJ (Maekawa et al., 2010), including 156 texts (5,815 sentences), with the discourse unit boundaries. We used the corpus for training a classifier to discourse unit boundary detection. As a feature set, we used lemmas and PoS of the head and functional morphemes appearing in the target and adjacent bunsetsu units. We also used lemmas of nouns and their PoS appearing in the dependency path from the bunsetsu unit in question to the end of the sentence. The dependency trees in the corpus were automatically analysed using a Japanese dependency parser, CaboCha. The results of a preliminary evaluation with 10-fold cross validation demonstrated that the classifier achieved reasonable performance (0.766 in F-score). We employed six revisers who have enough experience in English education into Japanese elementary schools.” Since

---

1. [http://www.cs.utah.edu/~hal/megam/](http://www.cs.utah.edu/~hal/megam/)
2. Bunsetsu is a basic unit of Japanese, consisting of at least one content word and more than zero functional words.
3. [https://code.google.com/p/caboche/](https://code.google.com/p/caboche/)
Early education of English in elementary schools is necessary. For example, the examination of bi-grams revealed that the segmentation of statement units and revisions of essays are essential in the examination. In this examination, we assume that the segmentation of statement units and revisions of essays are essential in the examination.

Because I think it is necessary to teach English at elementary school so that children can learn English before they feel difficulty in learning English in junior and senior high schools.

SU9 back1 DU9 私たち（がは）小学校中年生、純粋年とともに英語で繋げる機会があった。When I was in the middle and upper grades of elementary school, I had a chance to learn English.

SU1 elab1 DU1 それが（がは）小学校中年生における「授業の役割」ということ（がよ）は、授業の役割でお楽しそうに、授業で学んだことを発表しようというようにしてみたところ、自分でなく誰が強めるような時間になりがちである。Luckily, I enjoyed my English class in junior high school as well as in elementary school.

SU6 supp2 DU11 それ（がは）はつまり、私たちだけでなく中学受験（も）英語の役割が難しい。ものであることができた。

SU9 solv DU19 そのため短期に子どもが楽しんで受験に繋げることができれば、それが子どもの可能性を広げるということには繋がるのではないかだろうか。

Table 2: Example of a human revision

The indexes of the labels were excluded for this investigation.
viser needed to consider a logical relation for merging the adjacent sentences. We compared the merged sentences in the revised texts with the corresponding sentences in the original essays to find that the revisers tend to restore the original compound sentences. This counterintuitive result suggests that manual revision following the procedure in Section 2 does not effectively work for sentence reconstruction. Thus, there is a room of reconsideration on the before-revision representation for revisers.

4. Automatically evaluating text coherence

We evaluated text coherence using the original and revised texts on the task of pairwise information ordering, identifying a more coherent text. We employed the existing two models for evaluating text coherence proposed by Barzilay and Lapata (2008) and Iida and Tokunaga (2012). The former model (the entity-grid model) exploits the local coherence based on the local transition of discourse entities in a coherence chain with their grammatical roles, and the latter model (the coref_coherence model) uses the coherence score based on automatically detected coreference relations in a text. For creating an entity-grid model, we adopted the four grammatical role labels (wa (topic), ga (subj), oshi (obj/objj) and others). For learning these models, we used the NAIST Text Corpus (Iida et al., 2007), where coreference relations were manually annotated.

In our evaluation, we used a sample of 70 texts out of the 120 essays in which the sequential order of the discourse units in original and revised texts is not identical. We evaluated how correctly a model selects the revised text out of the original and revised texts. As a baseline model, we used a model that randomly selects from two given texts.

Table 5 shows the results of evaluating text coherence with our data set. For comparing them with the previous work (Iida and Tokunaga, 2012), the table also demonstrates the results using newspaper articles in the NAIST Text Corpus (Iida et al., 2007). Surprisingly, the two coherence models did not outperform the baseline when evaluating our data set, whereas the results of the previous work presented that these two models achieved significantly better performance than the baseline. The difference was caused by the different experimental setting between the two data sets. In the previous work using newspaper articles, the competing incoherent texts were generated by randomly reordering sentences in line with the experimental setting in Barzilay and Lapata (2008). In contrast, because our original essays were written by high school students, their coherence is obviously higher than the texts generated by random reordering. Eventually, given a pair of the original and revised essays, identifying a more coherent text becomes more difficult in comparison with the problem setting in the previous work.

In addition, newspaper articles, which were used as target texts in previous work, are written by professional writers to preserve high coherence by cohesively introducing discourse entities in a text. Therefore, the two coherence models relying on the local transition of discourse entities with grammatical roles relatively easily estimated the text coherence.

On the other hand, the discourse entities in the essays rarely cooccurred across these statement units. For example, word ‘English’ and phrase ‘elementary school’ are relatively frequently occurred over the text because they are related to the essay topic, early education of English in Japanese elementary schools. However, they rarely become centers but rather are used for modifying other words, e.g. English songs and in elementary school. Instead, the coherence of the essays is established by means of implicit logical relations between statement units, e.g. causal and elaboration relations. Therefore, estimating text coherence by the above models becomes more difficult compared to the case of newspaper articles.

5. Conclusion

This paper described building a corpus of manually revised texts which includes both before- and after-revision information. In order to create such corpus, we proposed a procedure for revising a text from a discourse perspective, consisting of dividing a text to discourse units, organising and reordering statement units and finally modifying referring and connective expressions, each of which imposes limits on freedom of revision. Following the procedure, six revisers who have enough experience in either teaching Japanese or scoring Japanese essays revised 120 Japanese essays written by Japanese native speakers. In this paper, we demonstrated some specific manual revisions frequently
C. Brockett, W. B. Dolan, and M. Gamon. 2006. Correcting ESL errors using phrasal SMT techniques. In Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics, pages 249–256.

R. Dale and A. Kilgarriff. 2010. Helping our own: Text massaging for computational linguistics as a new shared task. In Proceedings of the 6th International Natural Language Generation Conference, pages 261–265.

B. J. Grosz, A. K. Joshi, and S. Weinstein. 1995. Centering: A framework for modeling the local coherence of discourse. Computational Linguistics, 2(1):203–226.

R. Iida, M. Komachi, K. Inui, and Y. Matsumoto. 2007. Annotating a Japanese text corpus with predicate-argument and coreference relations. In Proceeding of the ACL Workshop ‘Linguistic Annotation Workshop’, pages 132–139.

M. Hermet and A. Désilets. 2009. Using first and second language models to correct preposition errors in second language authoring. In Proceedings of the 4th Workshop on Innovative Use of NLP for Building Educational Applications, pages 64–72.

R. Iida and T. Tokunaga. 2012. A metric for evaluating discourse coherence based on coreference resolution. In Proceedings of COLING 2012, pages 483–494.

W. C. Mann and S. A. Thompson. 1988. Rhetorical structure theory: Toward a functional theory of text organization. Text, 8(3):243–281.

I. Persing, A. Davis, and V. Ng. 2010. Modeling organization in student essays. In Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, pages 229–239.

M. D. Shermis and J. Burstein, editors. 2003. Automated Essay Scoring: A Cross Disciplinary Perspective.
Lawrence Erlbaum Associates.
Kei Usami. 2009. A study of psychometric characteristics of evaluation data on essay-type examination. Master’s thesis, The university of Tokyo.