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 Ôokayama, 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, organising 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-

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

Table 1: Statement unit labels

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 maker *wa* (topic) are off-topicalised by replacing with either *ga* (subj), *o* (obj) or *ni* (iobj) 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 $main_1$ and $elab_1$.

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 example, a reviser needs to explicitly edit a part of a sentence as ⟨A|B⟩ for replacing word A with word B in a sentence.

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

3. Manually revising Japanese essays

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[1] which judges whether a *bunsetsu*[2]-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*[3]. 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 either teaching Japanese or scoring Japanese essays. They were equally divided into two groups. The 120 essays were also equally divided into two data sets, and each reviser

---

[1] 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/cabocha/

suID	label	duID	revised results
SU4	the	DU7	〈 ただし |〉 小学校における英語の早期教育 〈 が | は 〉 必要である 〈 という |〉。←**Author's thesis should be placed at the beginning of a text**
			Early education of English in elementary schools is necessary.
SU3	main₁	DU5	〈 そのような意味では、|〉 中学校や高校で英語を学習して、「英語が難しい」と苦手意識を持ってしまう前に、小学校で「英語が楽しい」と思えるような教育をするの 〈 が | は 〉、むしろ必要なこと 〈| と思う 〉。←**Main idea is placed after author's thesis**
		DU6	日本語と同様に、相手とのコミュニケーションをとる手段として早期から英語に触れていれば、後になってから苦労して学ぶということに 〈| は 〉ならない。
			Because I think it is rather necessary to teach English at elementary school so that children have fun in learning English before they feel difficulty in learning English in junior and senior high schools.
SU0	back₁	DU0	私 〈 が | は 〉 小学校中学年、高学年のときに英語に触れる機会があった。
			When I was in the middle and upper grades of elementary school, I had a chance to learn to English.
SU1	elab₁	DU1	それ 〈 が | は 〉 中学校や高校における「英語の授業」という 〈。| よりは、〉
		DU2	遊びの感覚で楽しめる 〈。|〉
		DU3	ものであった。
			It was not a typical 'English class' in junior and senior high schools, rather it was an enjoyable play.
SU2	supp₁	DU4	例えば、英語の歌を歌ったり、朝の健康観察のときに先生と英語であいさつをしたり、英語を身近に感じることができるものであった 〈| と思う 〉。
			For example, we sang English songs, exchanged morning greetings in English in the classroom, which made English closer to us.
SU5	main₂	DU8	〈 前述の通り | ただし、〉 子どもたちが「楽しい」と思えるよう 〈 だ。| な 〉
		DU9	ものでない 〈。| と 〉
		DU10	〈| 英語の早期教育は 〉 全く意味がなく、むしろ逆効果になってしまう。
			However, if the experience is not enjoyable for children, early education does not make sense, rather it works negatively.
SU6	supp₂	DU11	私 〈 が | は 〉 幸運なことに、小学校だけでなく中学校で 〈| も 〉 英語の授業が「楽しい」ものであると思うことができた。
		DU12	それ 〈 が | は 〉 何よりまず、当時の ALT の先生のおかげであったかもしれない。
		DU13	というのも、彼女と私に 〈| は 〉 共通の趣味があり、よくそのことについて話したり、英語の授業で洋楽を聴いたり、英語で書かれたレシピを見ながらクッキーを作ったりと、私だけでなく誰が楽しめるような時間にしてもらえた 〈| からである 〉。
			Luckily, I enjoyed my English class in junior high school as well as in elementary school.
SU8	reb	DU18	「そんなに早くから子どもに英語を学ばせる必要 〈 が | は 〉 ない」という人 〈 が | も 〉 いるであろう 〈。| が、〉←**Rebuttal and its corresponding author's opinion are placed at the latter part of this text**
			Some would say that it's not needed to learn English at an early age. However,
SU9	solu	DU19	その早い時期に子どもが楽しんで英語に触れることができれば、それが子どもの可能性を広げるということに繋がるのではないだろうか。
			if children could enjoy English in the early stage, it would expand children's potential.
SU7	conc	DU14	英語に限ったことではない 〈。| が、〉
		DU15	何かを学ぶということにおいて、一番大切なの 〈 が | は 〉、学ぶ本人がどれぐらい意欲を持 〈 つ。|って 〉
		DU16	学べるかということである。
		DU17	「もっと知りたい」という意欲が何より本人の力を伸ばすのである。
			It's not just limited to teaching English. The most important thing is being motivated when they learn something.

The bold sentences stand for the comments by a human reviser with regards to the move of statement units. The numbering of duIDs denotes the order of the original text. The literal meaning in each discourse unit is sometimes difficult to translate into English preserving their nuance because they are not always a complete sentence. Thus, we instead translated the revised sentences in each statement unit into English.

Table 2: Example of a human revision

group revised a different data set following the procedure introduced in Section 2.

An example of a revised text is shown in Table 2, where 20 discourse units constitute 10 statement units. In this example, three types of revisions were performed. First, the 'thesis' (i.e. statement unit SU4) was placed at the beginning of the text. Second, the first 'main idea' (i.e. statement unit SU3) was swapped with its 'elaboration' (i.e. statement unit SU1). Finally, the 'conclusion' (i.e. statement unit SU7) was placed at the end of the text. These revisions contribute to improving the coherence of this text.

Given the original and revised texts, we can analyse general strategies of text revision by human revisers. To investigate what kinds of manual changes frequently occurred between the original and revised texts, we examined frequencies of the label bi-gram of two adjacent statement units[4] as shown in Table 3. Note that the statement unit labels were annotated to the revised version of essays. In this examination, we assume that the segmentation of statement units and their labels are identical in before- and after-revision texts. Thus, we sorted statement units in ascending order and then examined frequencies of the label bi-grams. Table 3 reveals notable tendencies of the revision. For example, the frequency of the bi-gram 'BOT (beginning of text)→the(sis)' drastically increased by the revision, conforming with our intuition that authors should first state their position on a

given topic.

Following the procedure introduced in Section 2, what expressions were modified are recorded by using the notation of '⟨original expression|revised expression⟩.' Thus, we can also analyse what expressions were modified through the revision. As a preliminary analysis, we investigated what modifications were frequently occurred in the revision. As a result, our manual analysis revealed that revisers typically employed the two types of modifications; the local and structural modifications. In the former modification, off-topic case makers *ga* (subj) and *o* (obj) are replaced with topic case marker *wa*, or appropriate connective expressions are inserted to an appropriate position in a sentence. On the other hand, as the structural modification, a sentence constituting a single discourse unit is merged with an adjacent logically related sentence to make a single sentence.

Table 4 shows the 20 most frequent modifications occurred in our current data set. It demonstrates that the most frequent modification is the topicalisation, i.e. 〈 が (subj)| は (topic)〉 due to the automatic pre-processing including off-topicalisation in discourse units.

Table 4 demonstrates that a reviser frequently employed the structural modifications as well. For example, two adjacent sentences are merged into a compound sentence with logical relations such as CONTRAST and CONDITION. Since the procedure in Section 2 decomposes a text into a series of discourse units with normalising their end phrases, a re-

---

[4]The indexes of the labels were excluded for this investigation.

su$_i$ \ su$_{i+1}$	EOT	main	elab	supp	the	back	conc	reb	solu	sugg	tran	pro	total
BOT	0 (0)	1 (-4)	0 (-2)	0 (-3)	85 (24)	7 (-16)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	4 (1)	97
main	10 (-7)	12 (-1)	87 (8)	45 (3)	3 (-7)	35 (10)	26 (5)	13 (-15)	0 (0)	1 (0)	10 (3)	1 (0)	243
elab	5 (0)	55 (-1)	27 (0)	68 (3)	4 (-3)	6 (-4)	13 (-4)	24 (9)	0 (0)	1 (-1)	7 (1)	0 (0)	210
supp	0 (-2)	57 (5)	42 (-5)	15 (2)	0 (-2)	9 (-1)	9 (-5)	17 (10)	0 (0)	0 (-1)	6 (-1)	0 (0)	155
the	4 (-3)	40 (3)	11 (3)	4 (-7)	1 (0)	34 (10)	1 (-2)	4 (-3)	2 (0)	0 (0)	5 (-2)	0 (0)	106
back	0 (0)	37 (1)	30 (1)	13 (2)	5 (-6)	5 (4)	2 (0)	7 (-1)	1 (0)	1 (0)	3 (-1)	0 (-1)	104
conc	75 (18)	2 (-4)	1 (-2)	0 (-2)	1 (-1)	0 (-1)	0 (0)	0 (-3)	0 (-1)	0 (-3)	0 (-1)	0 (0)	79
reb	0 (0)	7 (-1)	3 (-1)	1 (0)	1 (-4)	1 (0)	0 (-1)	8 (4)	36 (2)	20 (-2)	2 (1)	0 (0)	79
solu	1 (-3)	9 (1)	0 (-3)	2 (0)	0 (0)	0 (-2)	12 (3)	2 (0)	5 (1)	18 (2)	2 (1)	0 (0)	51
sugg	2 (-3)	6 (1)	4 (1)	4 (1)	1 (-2)	2 (-1)	13 (4)	0 (-2)	7 (-2)	5 (3)	2 (0)	0 (0)	46
tran	0 (0)	17 (0)	5 (0)	3 (0)	1 (0)	5 (0)	3 (0)	3 (1)	0 (0)	0 (0)	1 (-1)	0 (0)	38
pro	0 (0)	0 (0)	0 (0)	0 (0)	4 (0)	0 (0)	0 (0)	1 (0)	0 (0)	0 (0)	0 (0)	0 (0)	5

The number in parentheses is the change from the frequency in the original text. BOT and EOT stand for the beginning and end of a text respectively.

Table 3: Bi-gram frequency of statement units before and after revision

model	accuracy (%)	(cf. newspaper)
baseline	50.0	50.0
entity grid model	25.7	70.7
coref_coherence model	38.6	76.1

Table 5: Results of evaluating text coherence

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 coreference 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), *o*/*ni* (obj/iobj) 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

modification	type	description	freq.
⟨ が | は ⟩	LOCAL	replace a case marker with a topic marker to topicalise the argument.	3,111
⟨ が | も ⟩	LOCAL	add the meaning of *also* to the argument.	607
⟨ 。 |⟩	STRUCTURAL	merge adjacent sentences.	388
⟨| は ⟩	LOCAL	add a topic marker to topicalise the argument.	348
⟨| と思う ⟩	LOCAL	add a phrase (*I think*) to indicate author's attitude.	346
⟨| も ⟩	LOCAL	add the meaning of *also* to the argument.	242
⟨ 。 | が、 ⟩	STRUCTURAL	make a sentence a subordinate clause of the adjacent sentence with the CONTRAST relation.	209
⟨ 。 | と ⟩	STRUCTURAL	make a sentence a complement of the adjacent sentence.	158
⟨ る。 | り、 ⟩	STRUCTURAL	merge adjacent sentences.	155
⟨ 。 | と、 ⟩	STRUCTURAL	make a sentence a complement of the adjacent sentence.	123
⟨ る。 |、 ⟩	STRUCTURAL	merge adjacent sentences.	104
⟨ だ。 | な ⟩	STRUCTURAL	make a sentence an adnominal clause of the adjacent sentence.	96
⟨| からだ ⟩	LOCAL	add a conjunctive (*because*) to indicate the CAUSE relation between adjacent sentences.	91
⟨ る。 | れば、 ⟩	STRUCTURAL	make a sentence a subordinate clause of the adjacent sentence with the CONDITION relation.	83
⟨| と思います ⟩	LOCAL	add a phrase (*I think*) to indicate author's attitude.	81
⟨ る。 | て ⟩	STRUCTURAL	merge adjacent sentences.	81
⟨| しかし、 ⟩	LOCAL	add a conjunctive (*however*) to indicating the CONTRAST relation between adjacent sentences.	77
⟨ 。 | し、 ⟩	STRUCTURAL	merge adjacent sentences.	71
⟨ る。 | れば、 ⟩	STRUCTURAL	make a sentence a subordinate clause of the adjacent sentence with the CONDITION relation.	65
⟨ だ。 | で、 ⟩	STRUCTURAL	merge adjacent sentences.	60

Table 4: Top 20 frequent surface modifications

occurred between the original and revised texts, e.g. 'thesis' statements were frequently placed at the beginning of a text. We also evaluated 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 because the existing two models evaluated text coherence heavily relying on the cohesion of discourse entities, while the coherence of revised essays is mainly established by means of certain types of logical relations between statement units.

As future direction, we are planning to create an automatic text revision model based on this corpus, which would be useful for applications supporting human to write more informative and persuasive texts. In this paper, we decomposed a procedure of revising texts into the three steps, as shown in Section 2. In particular, the second step, organising and reordering statement units plays a key role among these three steps. To realise this step, it is important to employ techniques introduced in automated essay scoring, e.g. Persing et al. (2010), which automatically annotates sentence and paragraph labels with regard to the organisation of a text. We are planning to adopt such techniques to identify statement unit labels in essays and then tackle reordering problems based on a sequence of statement units with their labels.

6. References

R. Barzilay and M. Lapata. 2008. Modeling local coherence: An entity-based approach. *Computational Linguistics*, 34(1):1–34.

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*, 21(2):203–226.

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.

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.

C. Leacock, M. Chodorow, M. Gamon, and J. Tetreault. 2010. *Automated Grammatical Error Detection for Language Learners*. Synthesis Lectures on Human Language Technologies. Morgan Claypool.

K. Maekawa, M. Yamazaki, T. Maruyama, M. Yamaguchi, H. Ogura, W. Kashino, T. Ogiso, H. Koiso, and Y. Den. 2010. Design, compilation, and preliminary analyses of balanced corpus of contemporary written Japanese. In *Proceedings of the 8th International Conference on Language Resou rces and Evaluation (LREC 2010)*, pages 1483–1486.

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

Y. A. Park and R. Levy. 2011. Automated whole sentence grammar correction using a noisy channel model. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 934–944.

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.