

小論文推敲のための談話単位の並び換え

飯田 龍

徳永 健伸

東京工業大学 大学院情報理工学研究科

{ryu-i,take}@cl.cs.titech.ac.jp

1 はじめに

近年、自然言語処理の教育的な応用に関する研究が盛んに行われている。ここでは、ある言語の文法的な誤りの検出や自動訂正 [3]、小論文などの文章の自動採点 [7] など、幅広い応用への取り組みがなされている。本研究では、教育応用という文脈において、特に文章の談話レベルの推敲支援に取り組む。この研究の背景として、特に日本の教育システムにおいて文章を作成するカリキュラムが貧弱なこと、また理工学系の学生は一般的に大学入試において小論文記述が含まれないため、文章作成に関する指導を受ける機会が少ないことがあげられる。この問題を解決するために、計算機で体系的に検出できる談話レベルの不備をどのような観点でどう扱うかを考える必要がある。

一般に文章作成・推敲の支援を考える場合、大きく内容レベルの記述支援と表現レベルの記述支援が考えられる。内容レベルの支援では、ある話題について文章作成者が文章を記述する場合に、どのような内容を盛り込むべきかという情報の選択や、その内容に依存した論理構成をどのように決定するかを支援する。一方、表現レベルの支援では、内容レベルの段階ですでに確定した論理的構成に基づき、首尾一貫性や結束性の観点からそれらを文章中でどう表現するかを考える。内容レベルの記述支援を自然言語処理で達成するためには、与えられた課題に対して書くべき情報を適切に選択し、それらの間の論理的な関係をその課題に特有のドメイン知識や世界知識に基づいて推定する必要があるが、現状ではその問題を解くことは難しい。一方、表現レベルの記述・推敲支援については、これまで研究が進められてきたセンタリング理論 [4] などの談話理論から得られる知見や、小論文自動採点 [7] の分野で議論されてきた話題など、さまざまな談話的な特徴が利用できると思われる。このような背景から、本研究では表現レベルの推敲の問題に着目し、どのように推敲を支援可能かを考える。

これまでの我々の取り組みでは、態の自動選択 [11] や参照表現の自動生成 [9] などの談話的な特徴から文章をどう生成すべきかという研究に加え、実際に人間が記述した小論文を対象にどのような表現レベルの修正が可能なのか、また、それがどのような特徴に基づくものなのかを調査するために添削の専門家に制約付きで修正を行わせた [10]。本研究ではこの修正の結果を利用し、添削の専門家が文章の首尾一貫性を高めるために小論文中の談話単位（節など）をどのように並べ替えたのかを調査

し、その並び換えを計算機で実現可能かについて検討する。まず 2 節で収集したデータの概要を示し、次に 3 節で小論文データを対象にどのような並び換えが可能かを調査した結果について報告する。最後に、4 節でまとめと今後の展望について述べる。

2 小論文修正データの概要

本研究で利用する小論文データは、宇佐美 [8] が収集した小論文のうち、「小学校の授業における英語の早期教育は必要であるか否かについて、あなたの意見とその根拠が明確になるように、800 字以内で論述しなさい」という課題に対して高校生が作成した 120 編の小論文である。この小論文に対し、談話単位を言説構成素という論理的なまとまりにまとめあげ、その構成素単位で最適な順序となるよう構成素を再配置し、その結果起こる結束性レベルの不整合を局所的に修正するという制約付きの修正作業によって、もとの文章と比較してより首尾一貫性が高くなるよう修正された文章を得る。この作業を作業員の修正する多様性も調査するために、作業員を 6 名雇用し、人手修正作業を行わせた。120 編を 60 編ずつに分割し、各 60 編に対して異なる 3 名の作業員が修正を行った。この修正時には、Persing ら [6] が導入した構成に関するラベルを参考に導入した 11 の言説構成素のラベル (表 1) を各言説構成素にアノテーションすることで、並び換えを行うための手がかりを明示的に与えた¹。

具体的な人手修正の例を表 2 に示す。この人手修正の例では、文章が 9 つの言説構成素から成るとアノテーションされ、特に談話単位 (7) から成る言説構成素が書き手の主張であるため、文章頭に再配置され、それに付随してその主張を裏付ける理由が後続するよう並び換えられている。さらに、結論となる言説構成素 9 の前に、反論とその解決策である言説構成素 7 および 8 が配置されている。このように、言説構成素のラベルを考慮することで、人手修正作業員が明示的にその文章の論理的内容とそれに応じた構成を吟味した上で並び換えを行っていることがわかる。

人手修正作業の結果、作業対象となる 120 編に対し作業員 2 名がその記述内容の論理的破綻などを理由に作業を行わなかった 6 編を除き、354 編の修正結果を得た。この修正された小論文の特徴を調べるために、まず作業員ごとに 1 編に対し、平均いくつの言説構成素ラベルを付与したのかを調査した。この結果を表 3 に示す。この

¹人手修正作業の詳細は文献 [10] を参照されたい。

表 2: 人手修正の具体例

言説構成要素 ID	ラベル	談話単位 ID	修正結果
0	the	(7)	〈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〉 。
1	main1	(5)	〈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いう前に、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〉。
		(6)	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。
2	back1	(0)	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。
3	elab1	(1)	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、〉
		(2)	遊びの感覚で楽しめる〈。 〉
		(3)	ものであった。
4	supp1	(4)	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ができるものであった〈 と〉思う〉。
5	main2	(8)	〈前衛の通り ただし、)子どもたちが「楽しい」と思えるよう〈だ。 な〉
		(9)	ものでない〈。 と〉
		(10)	〈 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。
6	supp2	(11)	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。
		(12)	それ〈が は〉何よりも、当時の ALT の先生のおかげであったかもしれない。
		(13)	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もしてもらえた〈 からである〉。
7	reb	(18)	「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、) 8
8	solu	(19)	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とは思えないだろうか。
9	conc	(14)	英語に限ったことではない〈。 が、) 9
		(15)	何かを学ぶということにおいて、一番大切な〈が は〉、学ぶ本人がどれくらい意欲を持〈つ。 つて〉
		(16)	学べるかということである。
		(17)	「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。

表 1: 言説構成要素のラベル

ラベル	名前	説明
pro	prompt	著者に提示された問題を繰り返しているだけの内容
tran	transition	次の話題に移るための言い回しのみを含む (e.g. 「この理由として 2 つの理由があげられる」など)
the	thesis	著者の与えられた問題に関する立場を表明する内容 (e.g. 「早期教育に賛成である」など)
main	main idea	thesis を支持する主な理由や議論を述べた内容
elab	elaboration	main idea に関する説明や考えなどを含む内容。ただし、具体例は除く
supp	support	main idea や elaboration に関連する具体例
conc	conclusion	小論文の要約や結論に関連する内容
reb	rebuttal	thesis や main idea に反対する意見や具体例などを含む
solu	solution	rebuttal で示された疑問・問題を解消する内容
sugg	suggestion	rebuttal の解決方法を示す内容
back	background	main idea などの背景に相当する記述

表より、作業者は与えられた小論文に対し、著者の主張 (the) を平均的に 1 つ付与していることがわかる。また、主な理由 (main) やその詳細 (elab), 具体例 (supp) の間の揺れはあるものの、2 つから 3 つの主張に対する理由を付与するという論説の典型的な構造を検出していることがわかる。しかし、言説構成要素ラベルの談話単位ごとの作業者間の一致率を調査した結果 (表 4), 一致率の平均は 4 割に満たないことがわかった。人手修正の作業は、小論文に記述された内容のみから書き手の意図を推定し、その推定された意図を反映した形で修正が行われたが、この意図の読み取りや首尾一貫性を高める修正としてどこまで許容するかが作業者ごとに異なったため、このような結果になったと考えられる。

さらに、作業者が行った談話単位の並び換えについても、作業者ごとにその作業の結果が異なることがわかった。一例として、修正前の文章と人手修正後の文章における談話単位の順序を Kendall の順位相関係数で調査した結果を表 5 に示す。この表に示された作業者ごとの順

表 3: 小論文 1 編に対する言説構成要素ラベルの平均頻度

	A1	A2	A3	B1	B2	B3	平均
pro	0.05	0.04	0.02	0.03	0.00	0.02	0.03
tran	0.38	0.55	0.23	0.32	0.24	0.25	0.33
the	1.13	1.04	1.03	0.98	1.00	1.15	1.06
main	1.62	2.80	2.62	1.80	2.07	3.05	2.32
elab	2.68	1.43	2.18	2.72	3.88	2.53	2.58
supp	0.97	2.23	1.22	0.87	2.00	1.48	1.45
conc	0.62	1.02	0.78	0.80	0.97	0.53	0.78
reb	0.42	0.63	0.58	0.82	1.12	1.15	0.79
solu	0.45	0.43	0.43	0.43	0.59	0.68	0.50
sugg	0.08	0.29	0.18	0.27	0.26	0.45	0.25
back	0.58	0.79	1.53	1.10	0.55	0.43	0.83

A1~A3, B1~B3 がそれぞれ同一の 60 編に対して作業を行ったことを表す。

表 4: 談話単位ごとの言説構成要素ラベルの一致率

作業者の対	一致率 (平均)	一致率 (最大値)	一致率 (最小値)
A1-A2	0.337	0.719	0.079
A1-A3	0.383	0.878	0.091
A2-A3	0.343	0.829	0.061
B1-B2	0.380	0.857	0.033
B1-B3	0.376	0.833	0.071
B2-B3	0.386	1.000	0.036

位相関係数の平均を見てわかるように、作業者ごとにどの程度並び換えを行っているかが異なることがわかる。例えば、作業者 A2 と A3 では A2 がかなり積極的に順序を並び換えているのに対し、A3 はほとんど並び換えていないことがわかる。また、修正結果を作業者間で比較した順位相関の平均を表 6 に示すが、この結果からもわかるように、作業者の並び換えの傾向は、その書き手の意図の読み取りの結果と並び換えの積極性に依ってかなり異なる。

3 言説構成要素の並び換え

2 節で述べたように、作業者ごとに談話単位の並び換えの傾向が異なるため、同一の小論文であっても並び換えの結果が大きく異なることがわかった。ただし、書き換えの作業は言説構成要素のラベルに依存して行われるため、付与されたラベル系列の特徴を捉えることで一般的な書き換えの処理を達成できる可能性がある。そこで、本研究では、アノテーションされた言説構成要素の範囲と

表 5: 修正前と修正後の順位相関の平均

作業者	A1	A2	A3	B1	B2	B3
τ	0.819	0.672	0.907	0.758	0.692	0.849

τ は Kendall の順位相関係数を表す。

表 6: 作業者間の順位相関 (言説構成素単位の順序)

	A1-A2	A1-A3	A2-A3	B1-B2	B1-B3	B2-B3
τ	0.679	0.798	0.682	0.589	0.692	0.664

そこに付与されたラベルに応じた自動並び換えの処理を考える。この処理は小論文記述のオーサリングツールを使って文章を作成する場合に、内容のまとまりごとにラベルを付与された結果から、文章を生成した結果に対して、首尾一貫性が高くなるよう構成素を並び換える処理に相当すると考えられる。ただし、文献 [10] で示したように、本研究で対象とする小論文のデータは英語の早期教育に関する一般的な内容が記述されたものであるため、語彙連鎖に基づく首尾一貫性の推定 [1] がうまく機能しないことがわかっている。このため、ノイズとなることが予想される内容に関する語彙・統語的な情報は捨象し、言説構成素の並びの情報のみで並び換えがどの程度達成されるかを調査した。

3.1 Ranking SVM を用いた並び換え

並び換えを行う手法には、順列モデル [2] などさまざまな手法が存在するが、ここで捉えるべき特徴は「著者の主張は文章頭に配置する」という大きな移動をとまなう場合と、「主な理由はその詳細な情報よりも一般的には前に配置する」という局所的な並び換えが必要となる。このため、並び換える対象となる言説構成素が修正前にもど位置に書かれているか、また、どのような論理展開の上でその位置に出現したのかという情報を捉える必要がある。これらの特徴を明示的に捉えるため、本研究では言説構成素もとの順序をリランキングで並び換えるという手法を採用した。順序を学習するアルゴリズムとして Ranking SVM [5] を利用し、表 7 に示す素性を用いる²。この際、修正前の談話単位での出現順序、つまり分類対象が何番目の談話単位であるかという情報を素性として利用することで、もとの位置関係を考慮した学習を行う。

学習事例作成の具体例を表 2 に示した修正の具体例をもとに説明する。表 2 から事例のラベルと素性の抽出に必要な情報を抜粋したものを図 1 に示す。順位の列がそれぞれ SVM への入力となる順位であり、ラベルの列に示された言説構成素のラベルもとに素性を抽出する。例えば、順位が 3 番目の back に関しては $label_i$ が back, $label_{i-1}$ が main, $label_{i+1}$ が elab の素性となる。また、prev_labelset, next_labelset として 3 行目の前後に出現した言説構成素のラベルすべてを導入することで、back が記述された相対的な関係を導入できる。この結果得られたランカーを用い、修正前の言説構成素の順序をランカーが出力する順序に従って並び換える。

3.2 評価実験

2 節で示した小論文 354 編を利用し、10 分割交差検定で評価を行った。評価尺度として Kendall の順位相関

²言説構成素のラベルに付与されているインデックスの情報 (e.g. main1 の 1 など) は削除して利用した。

表 7: 素性

素性名	説明
$label_i$	分類対象となる i 番目の言説構成素のラベル (e.g. the, main)
prev_BOT	分類対象となる言説構成素が文章頭に出現するか否か
next_EOT	分類対象となる言説構成素が文章末に出現するか否か
$label_{i-1}$	分類対象となる言説構成素の一つ前のラベル (e.g. the, main)
$label_{i+1}$	分類対象となる言説構成素の一つ後のラベル (e.g. the, main)
prev_labelset	分類対象となる言説構成素より前に出現する言説構成素のラベル。素性値は出現回数
next_labelset	分類対象となる言説構成素より後に出現する言説構成素のラベル。素性値は出現回数
org_order	修正前の出現順序

順位	修正前の順序	ラベル
1	5	the
2	4	main
3	1	back
4	2	elab
5	3	supp
6	6	main
7	7	supp
8	9	reb
9	10	solu
10	8	conc

図 1: 表 2 からラベルと素性の抽出の例

係数を用い、ベースラインには修正前の小論文に関する順位相関係数の値を用いた。言説構成素を単位とした順位相関と談話単位を単位とした順位相関の 2 つの結果を表 8 に示す。この結果より、まず全事例に関して提案手法とベースラインを比較した場合、ベースラインと比較してリランキングを用いる提案手法の出力する順序がより修正後の順序に近い結果を得ていることがわかる。また、作業単位で評価した結果でも、作業者 A1 と B3 以外の 4 名に関する修正に関しては、提案手法に基づく並び換えの結果が修正前と比較して順位相関が向上した。これらの結果より、言説構成素のラベルとその順序のみを利用しても修正前と比較して順位相関に関しては修正後の順序に近づけることが可能であることがわかる。また、順位相関係数の分布がどのように変化したかについても調査した結果を表 9 に示す。この結果より、順位相関係数が負の値である事例を若干正の値に修正できていることがわかる。また、高い順位相関係数 (1.0~0.8) の事例数も増えていることがわかる。

実験で利用した 354 事例のうち、提案手法を用いて順位相関が向上した事例は 106 事例、一方、順位相関が低下した事例が 91 事例存在した。まず、順位相関が向上した 106 事例をすべて調査したところ、順位相関が向上する要因の多くは、「最初に著者が問いに対する主張を述べ、次にその理由を示し、さらにその詳細な根拠や具体例を述べ、必要に応じて反論やその解決策を述べたあとで、帰結を述べる」という小論文の典型的な構成となるように言説構成素を移動させる並び換えが起こったことで、順位相関が修正後の結果に近づいた。実際にどのような並び換えが起こったのかを調査した結果を表 10 に示す。この表より、例えば、著者の主張 (the) を文

表 8: 実験結果 (順位相関の平均)

	言説構成素		談話単位	
	ベースライン	提案手法	ベースライン	提案手法
全事例	0.761	0.779	0.784	0.789
A1	0.803	0.785	0.819	0.807
A2	0.646	0.700	0.672	0.705
A3	0.876	0.906	0.907	0.920
B1	0.720	0.758	0.758	0.760
B2	0.676	0.724	0.692	0.732
B3	0.834	0.795	0.849	0.806

表 9: 実験結果: 順位相関の分布

τ	ベースライン	提案手法
1.0~0.8	198	205
0.8~0.6	72	77
0.6~0.4	48	48
0.4~0.2	18	17
0.2~0.0	0	0
0.0~-0.2	5	1
-0.2~-0.4	1	0
-0.4~-0.6	0	0
-0.6~-0.8	0	0
-0.8~-1.0	0	0

頭に配置する, 理由 (main) を相対的に文章の前方に配置するといった並び換えが起こっていることがわかる. 一方, 順位相関が低下した場合についても調査を行った結果を表 11 に示す. 順位相関が低下した理由としては, 修正後でも小論文の典型的な構成とは異なった構成で文章が記述される場合である. 例えば, 典型的には著者の主張 (the) は文章頭に配置されるが, 下記の修正後の文章の冒頭では “back → the → main” の順序で記述されている.

back	今、若者における日本語の崩壊とも呼ぶべき現象が進行している。ら抜き言葉や KY 語といったものから、敬語の誤用まで実に多岐にわたっている。そんな中、小学校における英語の早期教育が必要だという声が上がっている。
the	しかし私はそれに意義を唱えたい。
main	なぜならば、語学には必ず母国語力がついてまわるのだ。

これは, 著者の主張 (the) とその理由 (main) の間に強い論理的なつながりがあるために, 背景 (back) をその間に挿入できないという理由から, この順序で構成されていると考えられる. しかし, 実際は main 内冒頭の「なぜならば」という接続表現は修正後に追記されたものであり, 修正前には存在しない. このため, この「私が早期教育に意義を唱えたい」と「語学には母国語力がついてまわる」という 2 つの命題の論理関係を命題の内容だけで判断する困難な問題が発生する. この種の問題は大規模な世界に関する知識とそれを用いた推論が必要となる非常に難しい問題であると考えられるが, 一方で, 小論文のテーマを限定することで, その知識と推論の過程を記述できる見込みもある. そのような閉じた世界を対象とした小論文の問題設定とそこで記述される可能性のある命題とその間の関係をあらかじめ規定しておくことで, この問題の一部は解決できると考えられる.

4 おわりに

本稿では, 文章推敲の支援の部分問題である言説構成素の並び換えの問題について, 既存研究 [10] で作成した小論文の人手修正データをもとに自動並び換えの可能性を検討した. 具体的には, 言説構成素の並び換えの問題をランキングの問題として扱い, 言説構成素

表 10: 正しく並び換えられた要因

要因	事例数
the を相対的に前方に配置	50
conc を文章末に配置	30
the を文章頭に配置	27
main を相対的に前方に配置	22
solu/sugg を相対的に後方に配置	13
reb を相対的に後方に配置	6
その他	6

表 11: 誤った並び換えの要因

要因	事例数
main は後方に配置したままにすべき	31
solu を後ろに移動しすぎ	21
2 つ目の the は後方に配置すべき	19
the の前に back を配置すべき	15
back は後方に配置すべき	6
その他	19

のラベルや修正前の位置情報を利用して並び換えることで修正後の並びに近づくことを示した. また, 並び換えられた結果を手で分析し, 順位相関が高くなった事例は小論文の典型的な構成へ並び換えが達成できていることがわかり, 一方, 順位相関が低くなった事例は言説構成素間の相対的な論理的なつながりを考慮して並び換える必要があることがわかった. 今後は, 本研究で提案した並び換えの手法と他の推敲支援の部分問題 (参照表現生成など) に関する手法を統合して自動推敲のシステムを作成し, 実際にユーザに利用させることで, 談話処理を利用した推敲支援の枠組み全体を評価したい.

謝辞

本研究は科研費若手研究 (A) 「談話解析技術に基づいた文章推敲支援」 (課題番号: 23680014) の支援を受けた. 記して謝意を表する.

参考文献

- [1] R. Barzilay and M. Lapata. Modeling local coherence: An entity-based approach. *Computational Linguistics*, Vol. 34, No. 1, pp. 1-34, 2008.
- [2] H. Chen, S.R.K. Branavan, R. Barzilay, and D. R. Karger. Global models of document structure using latent permutations. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 371-379, 2009.
- [3] R. Dale and A. Kilgarriff. Helping our own: Text massaging for computational linguistics as a new shared task. In *Proceedings of the 6th International Natural Language Generation Conference*, pp. 261-265, 2010.
- [4] B. J. Grosz, A. K. Joshi, and S. Weinstein. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, Vol. 21, No. 2, pp. 203-226, 1995.
- [5] T. Joachims. Optimizing search engines using clickthrough data. In *Proceedings of the ACM Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 133-142, 2002.
- [6] I. Persing, A. Davis, and V. Ng. Modeling organization in student essays. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pp. 229-239, 2010.
- [7] M. D. Shermis and J. Burstein, editors. *Automated Essay Scoring: A Cross Disciplinary Perspective*. Lawrence Erlbaum Associates, 2003.
- [8] 宇佐美慧. 小論文試験による評価データの心理計量学的性質の検討-制限字数の影響に焦点を当てて-. 修士論文, 東京大学, 2009.
- [9] 飯田龍, 徳永健伸. 日本語書き言葉を対象とした参照表現の自動省略-人間と機械処理の省略傾向の比較-. 情報処理学会自然言語処理研究会予稿集, NL-206-15, 2012.
- [10] 飯田龍, 徳永健伸. 談話レベルの推敲支援のための人手修正基準. 言語処理学会第 19 回年次大会発表論文集, pp. 830-833, 2013.
- [11] 飯田龍, 徳永健伸. 談話的な手がかりを利用した日本語の節の受動化. 言語処理学会第 17 回年次大会発表論文集, pp. 662-665, 2013.