

E-003

主題連鎖と文モダリティの分類に基づく小論文の自動採点  
Automatic Scoring of Short Essays based on Thematic Chain and Sentence Modality

藤田 彬<sup>†</sup> 田村直良<sup>‡</sup>  
Akira Fujita Naoyoshi Tamura

1. はじめに

本稿では、文間の主題連鎖と文モダリティの分類に基づいて日本語小論文を自動的に採点する手法について述べる。近年、無人でシステムとのインタラクションを進めていく形態が求められているが、その中で自由記述式問題の自動採点に対するニーズが高まっている。与えられた問題文に対して回答者が意見を述べる論説文を小論文と呼ぶが、小論文の自動採点システムは自由記述式問題の自動採点の中でも特にニーズが高まっている。

小論文の自動評価に関する先行研究は[1]のサーベイが詳しい。代表的なシステムとして、英文を対象にしたETSのe-raterなどが挙げられる[2]。また、和文を対象にしたシステムも、石岡らがe-raterを参考にJessというシステムを開発している[3]。しかし、Jessでは、接続表現のみを手掛かりにして文脈を把握するなど、筆者の主張についての論理の展開を評価の観点に加えることに対して十分に配慮されていないのが現状である。我々はこの状況に対し、主題連鎖を捉えることで筆者の主張についての論理の展開を把握して評価する手法を検討した[4]が、論理の展開の把握に関して精度に不十分な点がある。

そこで本研究では、論説文の文脈理解の精度向上に向けて、文章の結束性を支える主題の連鎖(主題連鎖)を捉えた上で、主題連鎖に含まれる各文における筆者の陳述態度(文モダリティ)を考慮する手法を用いて、「文章中である話題を提起し、論理的な誘導を行い、結論づける」という一連の論理の展開(論理展開)を捉える手法を検討する。また、論理展開に関する素性を用いて、文章を論理性について自動的に採点する手法を検討する。

2. 文モダリティの分類

2.1 意見文と叙述文

本研究における意見文と叙述文の定義について述べる。意見文は、文脈上で筆者の意見が述べられている文と定義する。また叙述文は、意見文ではない文と定義する。

2.2 文モダリティ・カテゴリ

本研究では、文脈を詳細に捉えるために、意見文を以下の2つの観点から4種類に分類する。

1) 特定陳述相手の有無

筆者の意見陳述が成立するために意見を訴えかける相手を**特定陳述相手**と呼ぶことにする。

- (1) 将来を見据えた教育を行なってほしい。
- (2) 小学校における英語の早期教育は必要である。

(1)には特定陳述相手があり、(2)には無い。(1)では「行なってほしい」と要求しているため要求する相手が存在しなければ意見陳述が成立しないからである。一方、(2)は陳述する相手が存在しなくとも意見陳述が成立する。

2) 特定対象の有無

筆者が陳述する意見が言及する対象を**特定対象**と呼ぶことにする。

(3) 初等教育は人間の基礎を作る重要な過程である。

(4) 国際人を育てるなら、まず日本語を身につけさせよ。

(3)には特定対象があり、(4)には無い。(3)では「初等教育」という特定対象に関して筆者の印象を述べている。一方、(4)には、特定対象が示されていない。

以上、1) 2) の2つの観点で意見文を表1のように4種類に分類する。また、叙述文を**カテゴリE**とする。

表1: 文モダリティ・カテゴリの分類

		特定陳述相手	
		有	無
特定対象	有	カテゴリ A	カテゴリ B
	無	カテゴリ C	カテゴリ D

3. 文間の主題連鎖

3.1 文の構造を捉えるモデル

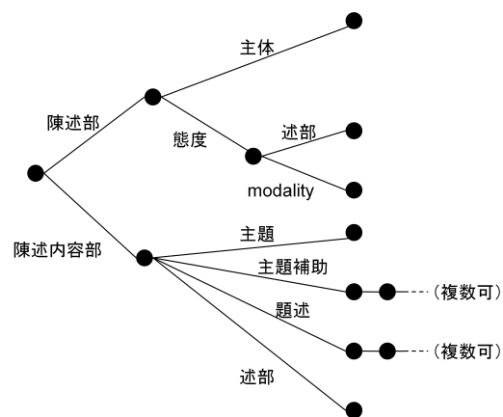


図1: 文の構造を捉えるモデル

図1に示したグラフ構造で、文の構造をモデル化する。以下で、このモデルについて述べる。

- 陳述部** 陳述内容に対する発話者についての記述
- 主体** 発話者
- 態度/述部** 主体の陳述態度に関する表現
- 態度/modality** 主体の陳述態度
- 陳述内容部** 陳述内容についての記述

<sup>†</sup> 横浜国立大学大学院環境情報学府  
<sup>‡</sup> 横浜国立大学大学院環境情報研究院

**主題** 文中で筆者が話題の中心としている対象  
**主題補助** 主題を修飾する語  
**題述** 主題と主題補助以外の語  
**述部** 用言

意見文では、陳述部に陳述内容に対する発話者について、陳述内容部に陳述内容について記述し、叙述文では陳述内容部に陳述内容部について記述する（陳述部は未定義）。

### 3.2 主題連鎖

本研究では以下の4種の主題連鎖関係に基づいて文間の連鎖構造を捉える。

#### ◆ 主題維持

ある文の主題（またはそれに類似する語）が直後の文の主題になっているか、または隣接する2文のうち後文の主題が省略されている2文間の関係

#### ◆ 主題変化

ある文の題述（またはそれに類似する語）が直後の文の主題になっている2文間の関係

#### ◆ 主題回復

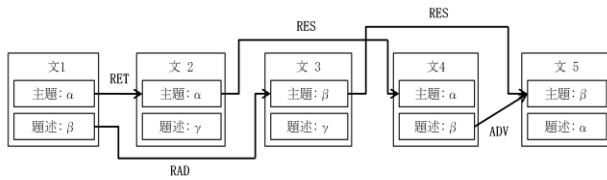
ある文の主題（またはそれに類似する語）が後（直後でない）の文の主題になっている関係

#### ◆ 遠隔主題変化

ある文の題述（またはそれに類似する語）が後（直後でない）の文の主題になっている関係

図2に以下の文章を例にとって主題連鎖の例を示す。

「1. 果たしてαはβなのか。 2. まずαはγである。 3. またβもγである。 4. よってαはβといえる。 5. またβはαともいえる。」



< RET : 主題維持, ADV : 主題変化, RES : 主題回復, RAD : 遠隔主題変化 >

図2: 主題連鎖の例

### 4. 文モダリティを含む文間の連鎖構造

本研究では、主題連鎖に基づいて捉える文間の連鎖構造と各文の文モダリティを複合したモデルを用いて、文章中の論理展開を把握する。

文間の連鎖構造は、論理が展開される流れ（パスと呼ぶ）を示すものである。また文モダリティは、文が陳述される時点で論理がどこまで誘導されているかを示すものである。例えば、論理展開が開始される文、すなわち話題が提起される文は、文モダリティが読み手に話題を投げかける種類の文モダリティ(カテゴリ A, カテゴリ C)で、パス上において始点に位置する文である。

このように、各文についてパス上の位置とモダリティを捉えることで文章の論理展開を把握することができる。そこで本研究では、文間の主題連鎖と連鎖する文のモダリティを同時に捉える以下のようなモデルを提案する。

$$m_1(X_1, Y_1)m_2(X_2, Y_2)$$

$m_1$ はモダリティ、 $X_1$ は主題、 $Y_1$ は題述をそれぞれ示すものである。このモデルによって、主題連鎖は以下のように定式化できる。

**主題維持**  $m(t, r)m'(t, r')$   
**主題変化**  $m(t, \{\dots, t', \dots\})m'(t', r')$   
**主題回復**  $m(t, r) \dots m'(t, r')$   
**遠隔主題変化**  $m(t, \{\dots, t', \dots\}) \dots m'(t, r')$

$$\begin{cases} t, t': \text{主題} \\ r, r': \text{題述} \\ m, m': \text{文モダリティ・カテゴリ} \\ m, m' \in \{A, B, C, D, E\} \end{cases}$$

このモデルを用いて図2の例を表すと、以下のように表される。

$$A(\alpha, \beta)E(\alpha, \gamma)E(\beta, \gamma)B(\alpha, \beta)B(\beta, \alpha)$$

### 5. 小論文自動採点システム

図3に本研究で提案する小論文の自動採点システムのフローを示した。文モダリティは、[5]で用いられている決定木学習に基づく手法で決定する。自動採点モデルには、重回帰分析を用いる。重回帰分析を行なう際に事例毎の目的変数となる値は、被験者により書かれた文章を専門家が採点したスコアを用いる。文章は605編あり、筆者である被験者は全員が高校生である。スコアには、論理性に関する5種の観点「簡潔」「明確」「構成」「一貫」「説得」がある。専門家は4名おり、それぞれ1~5点の整数値で採点した。最高点と最低点を除外したスコアの平均を専門家の採点結果としてまとめる。ただし、専門家全員が異なるスコアを出した場合は、対象となる文章は単一のモデルでの採点が困難であると判断し、重回帰分析の対象から除外する。また、重回帰分析の説明変数として用いる素性は118個ある。各素性の詳細については付録で述べる。

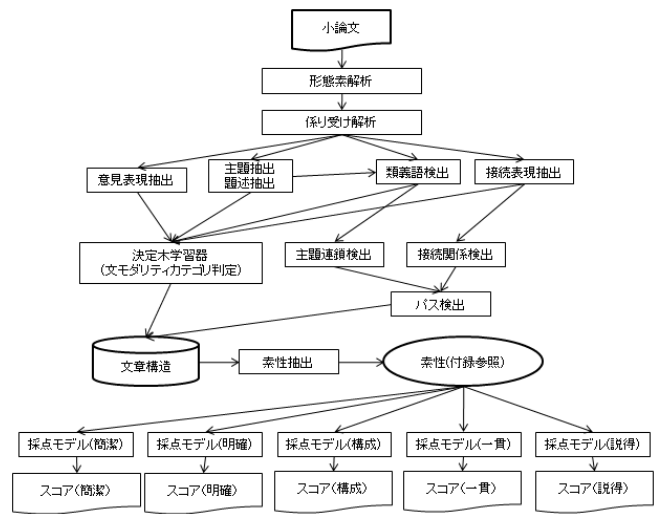


図3: 提案する自動採点システムのフロー

### 6. 実験

#### 6.1 実験 (連鎖構造)

提案モデルで実際的小論文の論理展開を捉える実験を以下の手順で行なった。

1. 被験者の高校生が書いた小論文 605 編を用い、それぞれを専門家 4 名が採点した。これらの文章を以下のようにクラス分けした: ①得点順に 5 段階にクラス分けをする, ②クラスの文章数の分布が正規分布になるようにクラスの境界を調整する。最も高い評価を受けた

文章を含むクラスを最高クラス、最も低い評価を受けたクラスを最低クラスと呼ぶことにする。

2. 最高クラスと最低クラスに属する文章について採点者4名の間での採点結果の標準偏差を測定し、標準偏差が小さい文章を各クラス20編ずつ抽出した。
3. 計40編の文章に含まれる主題連鎖と文モダリティを手動で判定した結果から、提案モデルで捉えられる主題連鎖と  $m_1(X_1, Y_1)m_2(X_2, Y_2)$  の組合せパターン (以下、展開パターン) について、各クラスでの出現頻度を求めた。

表2に最高クラスでの展開パターンの出現頻度、表3に最低クラスでの出現頻度を示す。両クラスの平均値の差の検定(t検定)を行なったところ、有意水準5%において有意差があると判定された ( $P(T \leq t)$  両側の値が 0.00427)。

## 6.2 考察 (連鎖構造)

$A(t, r) \dots B(t, r')$  は、読み手に投げかける形で問題提起を行い、提起した問題を確認しながら結論を述べる展開パターンである。この  $A(t, r) \dots B(t, r')$  は、最高クラスに出現する反面、最低クラスには出現しない。このことから、明確な問題提起を行なった上で何に対する結論であるかを提起段階に遡って結論を示す論理展開になっている場合、読み手に好印象を与える要因となりうるということがわかる。

$B(t, r) \dots B(t, r')$  は、論理展開中の離れた文間で意見陳述する特定対象が一致していることを示す展開パターンである。これは、全体的に話題が一貫していることが表れる展開パターンと解釈することができる。この  $B(t, r) \dots B(t, r')$  が各クラス内で出現する割合は、最高クラスで 11.1% である事に比べ、最低クラスでは 4.00% と少ない。このことから、読み手に与える印象が良い文章では、より意見陳述の特定対象が一定した一貫性のある論理展開になっていることがわかる。

$B(t, \{ \dots, t', \dots \}) E(t', r')$  は、ある特定対象に対する意見陳述の後に、題述の内容に関する事実確認を行う展開パターンである。これは、意見陳述の後に判断の根拠を示す展開パターンと解釈することができる。この  $B(t, \{ \dots, t', \dots \}) E(t', r')$  が各クラス内で出現する割合は、最高クラスで 6.75% である事に比べ、最低クラスでは 0.80% と少ない。このことから、読み手に与える印象が良くない文章では、事実確認による判断の根拠を示した意見陳述が少ないことがわかる。

$E(\dots) E(\dots)$  の主題連鎖は全て、事実確認をする文間に連鎖がある展開パターンである。この展開パターンは、事実陳述が連続していることを示すものである。意見陳述を目的とする論説文における事実陳述は意見陳述の判断根拠の役割を果たすものであるため、冗長な事実陳述は望ましくない。例えば、参考資料の内容の羅列が大部分を占める文章にこの展開パターンが多く見受けられる。 $E(\dots) E(\dots)$  の主題連鎖の各クラス内で出現する割合は、最高クラスで 9.92% であることに比べ、最低クラスでは 19.2% と多い。このことから、連続した事実陳述は度合いによって読み手に良くない印象を与える要因の一つとなりうるということがわかる。

## 6.3 実験 (自動採点)

次に、提案する自動採点システムの評価実験を以下の手順で行なった。

1. 被験者 (高校生) が書いた 601 編の小論文それぞれについて 118 個の素性を抽出する。各小論文には、専門家 4 名による論理性に関する 5 種の評価観点に関しての採点結果の合意をとったスコアが与えられる (合意のとり方は前述の通り)。5 つの評価観点には「簡潔」「明確」「構成」「一貫」「説得」がある。
2. 専門家により各小論文に与えられたスコアを目的変数、提案システムにより各小論文から抽出した素性を説明変数として、重回帰分析を行う。予測値と実測値の残差平均<sup>1)</sup>と、残差の絶対値が 0.5 以内の場合に一致したとみなした上での一致率で、重回帰分析の結果を検定する。検定は 5 分割交差検定である。
3. AIC に基づくステップワイズ法で説明変数を選択する。本実験の結果として、予測値と実測値の残差平均を表 4 に、残差絶対値が 0.5 以内の場合に一致したとみなした上での一致率を表 5 に、それぞれ 5 種の評価観点毎に示す。

## 6.4 考察 (自動採点)

表 4 を参照すると、「一貫」に関する残差平均が低いことがわかる。このことから、提案するシステムで素性として採用しているパスに関する素性が文章の一貫性の採点に有効であることがわかる。提案するシステムでは、主題連鎖を基にしてパスを捉え、パスの長さの最大値やパスの総数などの素性を用いて採点するが、ステップワイズ法においてこれらのパスに関する素性は全て選択されている。このことから、パスに関する素性が文章の一貫性の採点に有効であることがわかる。一方、表 5 を参照すると「一貫」に関する一致率が低いことがわかる。このことから、平均的に残差は小さいものの極端に残差の大きい事例があることがわかる。そのような事例には、提案するシステムで捉えられないパスが存在することが考えられる。

## 7. おわりに

本稿では、文間の主題連鎖と文モダリティの分類に基づいて日本語小論文を自動的に採点する手法について述べた。

提案する自動採点システムは、2 つのモジュールを含み、1 つ目のモジュールでは文モダリティと主題連鎖を複合したモデルで文章構造を捉え、2 つ目のモジュールでは文章構造上の素性を用いた重回帰モデルで文章を論理性に関して自動的に採点する。実験の結果、提案するシステムが文章の論理性に関する自動採点に有用であることが分かった。特に、論理性の中でも文章の一貫性に関しての採点に有用であることが分かった。

今後の課題として、主題連鎖の検出精度とパスの検出精度の向上が挙げられる。また、文モダリティ・カテゴリの分類の細分化も挙げられる。現状での文モダリティ・カテゴリの分類を、文章中における文の役割を表現出来るさらに細かい分類にすることで、筆者の意図した論理展開をさらに高い精度で把握することが可能になると考えられる。

1)  $\frac{\sum_{i=1}^n \sqrt{(y_i - f_i)^2}}{n}$  (残差平均)

$y_i$  は実測値、 $f_i$  は予測値、 $n$  は小論文の数をそれぞれ示す。

表 2: 最高クラスでの展開パターンの出現頻度

	$\{m, m'\}$									
	AA	AB	AC	AD	AE	BA	BB	BC	BD	BE
維持	0	2	0	0	1	3	31	0	4	5
変化	0	1	0	1	0	2	23	0	4	17
回復	0	4	0	0	0	3	28	0	2	5
遠変	0	1	0	0	0	1	7	0	1	5
	CA	CB	CC	CD	CE	DA	DB	DC	DD	DE
維持	0	0	0	0	2	0	1	2	0	1
変化	0	0	0	0	0	2	0	0	0	3
回復	0	0	0	0	1	0	0	0	0	1
遠変	0	0	0	0	0	0	0	0	1	0
	EA	EB	EC	ED	EE					
維持	0	16	1	2	13					
変化	3	18	0	2	8					
回復	0	13	0	2	3					
遠変	0	5	0	0	1					

表 3: 最低クラスでの展開パターンの出現頻度

	$\{m, m'\}$									
	AA	AB	AC	AD	AE	BA	BB	BC	BD	BE
維持	0	1	0	0	1	2	16	0	1	8
変化	0	0	0	1	0	0	5	1	3	1
回復	0	0	0	0	0	0	5	0	2	3
遠変	0	0	0	0	0	0	0	0	0	1
	CA	CB	CC	CD	CE	DA	DB	DC	DD	DE
維持	0	0	1	1	1	0	2	0	2	0
変化	0	0	1	0	0	0	2	0	3	1
回復	0	0	0	0	0	0	2	0	1	0
遠変	0	0	0	0	0	0	2	0	0	1
	EA	EB	EC	ED	EE					
維持	2	8	0	1	10					
変化	0	8	0	4	7					
回復	0	5	0	1	5					
遠変	1	0	0	0	2					

表 4: 専門家の採点結果との残差平均

簡潔	明確	構成	一貫	説得
0.638	0.672	0.688	0.598	0.684

表 5: 専門家の採点結果との一致率

簡潔	明確	構成	一貫	説得
0.452	0.426	0.409	0.391	0.428

付録: 提案するシステムで用いられる素性

ID	Features
1 to 100	主題連鎖のある文間での文モダリティの展開パターン(100種)の出現頻度
101	意見文の数
102	意見の正当性: $\sum_{i=1}^A D_i$ Aは意見文の数, $D_i$ は意見文に至るパスの長さ
103	話題の発展性: 意見文間のパスの最長値
104	結論との結束性: 結論段落とリンクがある文の数
105	文章全体の結束性: リンクの総数
106	段落間の関連性: 異なる段落間のリンクの個数
107	段落内の結束性: $\sum_{n=1}^P \frac{L_n}{S_n}$ $L_n = n$ 段落内のリンクの個数 $S_n = n$ 段落の文の数
108 to 112	各文モダリティ・カテゴリの出現頻度
113	[6]で用いられている接続関係「論理」の出現頻度
114	[6]で用いられている接続関係「論理」の出現頻度
115	[6]で用いられている接続関係「論理」の出現頻度
116	[6]に紹介がある明示的な照応関係についての素性
117	段落の数
118	文の数

参考文献

- [1] 石岡恒憲, “小論文およびエッセイの自動評価採点における研究動向”, 人工知能学会誌, Vol.23, No.1, pp.17-24(2008)
- [2] Yigal Attali, Jill Burstein, “Automated essay scoring with e-rater v.2”, *Journal of Technology, Learning and Assessment*, 4. Retrieved June 22(2007)
- [3] 石岡恒憲, 亀田雅之, “コンピュータによる小論文の自動採点システム jess の試作”, 計算機統計学, Vol.16, No.1, pp.3-18(2003)
- [4] 藤田彬, 田村直良, “文章構造解析に基づく小論文の自動評価”, FIT 2008(2008)
- [5] 藤田彬, 鈴木春菜, 田村直良, “文脈理解のための文モダリティの分類と自動判定”. 情報処理学会研究報告, 2009-NL-189(2009)
- [6] 横野光, 奥村学, “テキストの結束性判定のための entity grid モデルの素性の検討”, 情報処理学会研究報告, 2009-NL-189(2009)