

文章作成・推敲教育に向けた詳細なアウトラインの仕様設計と 修辞構造情報付与の試み

藤田 篤 柏野 和佳子 大塚 裕子 富永 敦子 椿本 弥生
情報通信研究機構 国立国語研究所 公立はこだて未来大学

1 はじめに

物事を論理的, 多角的に捉えて表現・伝達する力(論理的思考力)は, 我が国では, 学士課程教育において習得すべき汎用的技能と位置付けられている [2]. この答申と前後して, 多くの高等教育機関において, 文章作成・推敲技術を習得させることを目的としたカリキュラムが実施されるようになってきた. 我々も, 高等教育機関における文章作成・推敲教育の方法論について, 実践を含めて検討している [6, 9].

本稿では, 我々が設計・検証を進めている, 論理的な文章を作成するためのツールについて述べる. まず, 文章に含める内容を網羅的・構造的に整理するための, 従来よりも詳細なアウトライン(以下, **詳細化アウトライン**)の仕様を示す. さらに, 文章構成要素間の論理関係の吟味を容易にするための, 修辞構造理論 [7] に基づく関係の体系と, 詳細化アウトラインに対してそのような関係を(人間の手で)付与する手続きについて述べる. 小規模ではあるが, 実際に詳細化アウトラインを作成し, 論理関係を付与した. それらを通じて得られた知見についても報告する.

2 詳細化アウトライン

2.1 アウトライン上での内容の網羅・構造化

文章の著者は, 論理的かつ読みやすい文章を作成するために推敲を重ねる. 推敲時には, 内容の論理性や構成(what-to-say)と言語表現を含む情報の提示方法(how-to-say)の両方の問題を解消する必要がある. しかし, 書き下された文章の推敲においては, 表現の統一, 非文法的あるいはぎこちない言語表現の修正などの校正, 局所的な表現の書き換えなど, how-to-say の問題の修正に留まってしまう傾向にある [3, 5]. 我々は, 書き下された文章には論理構造が明示されていないことが原因の一つであると考え.

一方, 文章作成に関する指南本等が述べるように, 論理的な文章を作成する上でのアウトラインの重要性は誰もが認めるところであろう. ただし, これらの文献においても, アウトラインの具体例としては章立てに相当するものしか示されておらず, パラグラフライティングやいわゆる文法として学ぶ技術との間に隔たりが

ある. 内容を十分に列挙・整理しないまま文章を書き始めると, 必然的に新たな事柄を追加することになるが, それを適切な位置に配置できなかつたり, そのような問題を推敲時に見逃したりする可能性が高く, 結果的に文章の論理性を損なってしまう.

文章作成・推敲時における what-to-say のレベルの吟味を容易にするために, 我々は, 最終的な文章に含める内容を網羅・構造化したアウトライン(詳細化アウトライン)の作成について検討している [6, 9]. 産出した文章の長さ等に応じて論述の細かさや話題の広がりとは異なりうるが, これは詳細化アウトラインの木構造を枝刈りして制御すること [4] に見立てる.

2.2 仕様の設計と洗練

加藤 [6] は, 学習者の計算機リテラシおよびデータ処理の利便性を考慮し, スプレッドシートを用いて詳細化アウトラインを作成することを提案した(図 1). そして, 3 件の小論文の模範解答(974, 811, 781 字)から詳細化アウトライン(最終的な構成要素数は 42, 32, 34)を作成し, 詳細化アウトラインが満たすべき要件の列挙と詳細化アウトラインの修正を繰り返し, 次に示す 4 個の大項目, 17 個の小項目にまとめた.

- (a) **内容に関する要件 (5 項目):** 自分の主張を明確にする, 主張の根拠を必要十分に含める, etc.
- (b) **構造に関する要件 (2 項目):** 明示的に段差げしてアウトラインを階層構造化する, 各行には 1 要素のみ配置する.
- (c) **個々の要素に関する要件 (6 項目):** 1 つの要素は 1 つの文または名詞句で書く, 接続詞や指示語を含めない, etc.
- (d) **リファレンスに関する要件 (4 項目):** リファレンスは書籍, 学術論文, 新聞記事, Web ページのいずれかで, かつ実在するものに限る, etc.

我々は, この仕様を, 新たなサンプルに基づいて修正した. まず, 言語学のバックグラウンドを持つ 3 名に, 上記の仕様を満たす詳細化アウトラインの作成を依頼した. その際, 最終的な文章は 800~1000 字の意見文とし, 大学入試小論文等で用いられるような 10 個のテーマを与えた. これまでに 21 件の詳細化アウトラ

要素 ID	段落見出し	第 1 層	第 2 層	第 3 層	第 4 層
12	発生時の対処				
13		時間との戦い			
14			速やかに行動するための指針の必要性		
15				フォーマット化された質問票	
16					定量的・効率的な状況把握のため
17				疫学的・統計的数な技術情報	
18					仮説検証のため

図 1: 「食品媒介疾患」について作成された詳細化アウトラインの一部。

インを得たが、空行・空列、構成要素内の取り消し線、未参照のリファレンスなどの形式上の不備が多数存在していた。これらを修正した結果(以下、**FY2013 データ**)、構成要素数は 26~57 (平均 41) となった。構成要素数はテーマや論じる内容に依存する。ただし、仕様(b)の「1つの要素は1つの文または名詞句で書く」ということが順守できていない場合や、逆に記述が不十分であるため、著者自身の意見、他者の意見の引用、客観的事実のいずれであるかを、我々読者が判断できない構成要素も見受けられた。

上記の問題を緩和するため、またトップダウンな情報整理の重要性 [8] を考慮し、次のような項目を新たに仕様に含め、4 個の大項目、22 個の小項目 ((a) 4 項目、(b) 6 項目、(c) 9 項目、(d) 3 項目) にまとめた¹。

- (b) **構造に関する要件:** 主要な内容を表す要素を左の方の列に、付属的な内容を表す要素を右の方の列に配置する。その際、まずは主要な内容の真偽・構造を十分に整理する(全体から部分への展開 [8])。
- (c) **個々の要素に関する要件:** 付属的な内容を新たに記述する際、主要な内容を表す要素全体と関連していなければ、後者を複数の要素に分解する。

3 修辞構造情報の付与方法

詳細化アウトライン上で構成要素間の論理関係を学習者に明確に意識させる手段として、我々は、修辞構造理論 (Rhetorical Structure Theory, RST) [7] に基づく関係を付与する/させることを検討している。

Carlson ら [1] は、RST の原著 [7] における修辞関係(以下ではこれらも論理関係と呼ぶ)を 78 種類(大分類としては 16 種類)に拡張し、英語の文章に対して修辞構造情報を付与した。これに対して、詳細化アウトラインでは談話要素が明示的かつ階層的に与えられている。また、学習者の教材としての用途を想定すると、論理関係の特定にかかる人的コスト、および学習者の到

¹紙面の都合上ここでは抜粋のみを示す。詳細は次の文書を参照されたい。 <http://paraphrasing.org/~fujita/FLIP/outline.pdf>

達目標の観点で、Carlson らの体系よりも粗い分類を必要とする。さらに、Carlson らの体系には後述の 2 つの問題がある。これらをふまえ、加藤 [6] は、論理関係の体系、および詳細化アウトラインに対して修辞構造情報を付与する手続きを改めて設計した。

3.1 関係の体系

加藤はまず、Carlson らの体系における次の 2 つの問題を解決し、改良 Carlson 体系を作成した。

問題 1. 単核関係と多核関係の混在: Carlson らの体系では、関係の種類を特定した後に、類似する単核関係²と多核関係³を区別する場合がある。加藤はこれらを最初に区別することとし、所与の 2 つの談話要素を単核関係、多核関係、関係なしのいずれかに分類するための言語テスト⁴も定めた。

問題 2. 関係の定義の方向の混在: Carlson らの体系における単核関係の中には、Nucleus に対する Satellite の役割と Satellite に対する Nucleus の役割が混在している。我々は、関係の種類を前者に統一した。また、文献 [4] にならい修辞構造を依存構造で表すことにし、次のように形式を定めた。

- Satellite が Nucleus に係る
- 多核関係は後方の要素が前方の要素に係る 2 項関係を連ねて表す

加藤は次に、FY2013 データとは別の自作した 3 件の詳細化アウトライン(構成要素数は 38, 38, 45)に対して、改良 Carlson 体系に基づく関係を、様々な可能性を考慮して付与した。そして、実際に観察された関係に基づいて次のように関係の種類を限定・体系化した。

取捨選択: 改良 Carlson 体系の関係のうち、サンプルにおいて観察されなかった関係を除外した。新たなサンプルにおいてそのような関係が観察された場

²中心的な要素(核, Nucleus)と、それを修飾する要素(衛星, Satellite)との間の関係。

³2 つ以上の要素が等しく重要である(等位の)関係。

⁴いずれか一方を削除した場合にもう一方のみでも元の 2 要素の持つ主たる内容が伝わるか否か、**削除テスト**と呼ぶ。

表 1: 単核関係の一覧 (16 種類, 5 つの大分類).

決定木を含む詳細は次の文書を参照されたい. <http://paraphrasing.org/~fujita/FLIP/DT-mononucleus.pdf>

大分類	関係を認める条件	具体的な関係の種類
D1. 著者の解釈/意見	Satellite と Nucleus の少なくとも一報が著者の解釈や意見を表す	証拠, 理由づけ, 評価
D2. 因果関係	Satellite と Nucleus の間には因果関係がある	原因, 結果
D3. 条件	Satellite が満たされない場合 Nucleus は成立しない	条件, 使用可能性
D4. 情報の明確化	Satellite は Nucleus の位置付けや内容を明確にする	言い換え, 要約, 比較
D5. 情報付加	Satellite は Nucleus に関する新しい情報を付加する	方法, 目的, 背景, 例示 詳細化, 帰属

表 2: 多核関係の一覧 (4 種類, 3 つの大分類).

決定木を含む詳細は次の文書を参照されたい. <http://paraphrasing.org/~fujita/FLIP/DT-multinucleus.pdf>

大分類	関係を認める条件	具体的な関係の種類
P1. 対照	2 つの要素が対照のことを述べている	対照
P2. 時間順序	2 つの要素が連続して起こることを述べている	継起, 逆継起
P3. 並列構造	2 つの要素が並列構造をなす	リスト

合は「その他」という関係を付与しておき, 必要に応じて分類する.

階層化: 元の Carlson の体系における 16 種類の中分類をまたがないように, 関係の種類を 2 つずつボトムアップに統合した. その際, 2 つの関係 (群) を区別するための条件を記した. 関係の階層化の際に与えた条件を繋げ, 単核関係, 多核関係の各々に関する決定木を得た.

以上が加藤 [6] の成果の概要である. 今回は, 2.2 節で得た FY2013 データのうち 3 件 (4 節表 3 の文書 D, E, F) に対して, 第 1, 第 2 著者が論理関係を付与し, それに基づいて論理関係間の優先順位および分類の条件を再調整した. こうして得られた関係の一覧を表 1, 表 2 に示す. 関係を同定するための決定木については, 各表に示す文書を参照されたい.

3.2 論理関係を付与する手続き

3.1 節では, 詳細化アウトラインに対して, 特定の手続きを定めずに論理関係を付与した. 詳細化アウトラインの仕様 (b), (c) に従って, ある程度適切に階層構造を与えることができていたため, 多くの要素が親要素⁵または兄弟要素⁶に属していた. このことから, 詳細化アウトライン上の構造に従ってボトムアップに論理関係を付与できると考えた. 一方, 子要素⁷や, 親子兄弟以外の要素に係る要素もあった. 例えば, 図 2 に示すように, 図 1 の要素 13 は, 子要素である要素 14 に対する「理由づけ」を表す.

以上をふまえ, 加藤 [6] は, 所与の詳細化アウトライン全体に対する修辞構造情報を付与する手続きを次の 4 ステップにまとめた.

⁵着目している要素 X の 1 段上の層にある要素の中で, X よりも前方にあり, X の最も近くにある要素. 図 1 を例に上げると, 要素 14 に対する要素 13, 要素 17 に対する要素 14 など.

⁶着目している要素と同じ親要素を持つ要素.

⁷着目している要素を親要素とする要素.

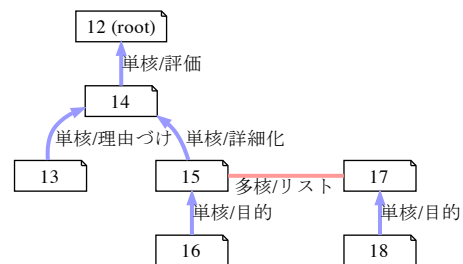


図 2: 図 1 のアウトラインに対する修辞構造情報.

ステップ 1. メタ要素の除外: 「背景」, 「まとめ」, 「考察」, 「例」のような, それ自体が内容を含まない要素を, 関係を付与する対象から除外する.

ステップ 2. 親子兄弟要素間の関係の同定: 個々の要素について, その親子兄弟要素の各々との関係の有無および種類を削除テスト, 関係の決定木によって判定し, その中で最も適切なものを選択する.

ステップ 3. 係り先が未定の要素の係り先の同定: 親子兄弟要素の中に係り先が存在しない要素について, その他の要素との関係の有無および種類を判定し, 最も適切なものを選択する.

ステップ 4. 全体の見直し: ボトムアップに定めた構造を, 改めてトップダウンな観点から見直す. 構造または関係を訂正する場合は, 当該 2 要素および新しく係り先と判定された要素の各々について, それを係り先とする要素との関係の適否も確認する.

我々は, アウトラインの仕様および論理関係の体系の再整理に合わせて, この手順も見なおした. 表現の統一等を行ったが, 大きな変更は必要ではなかった⁸.

4 修辞構造情報の付与実験

FY2013 データは質的に多様であり, 各基準を作成する際に参照したものと比べて質が低いものも含まれる.

⁸詳細は次の文書を参照されたい.

<http://paraphrasing.org/~fujita/FLIP/rst4outline.pdf>

表 3: 修辞構造情報の付与結果.

ID	要素数	4人一致						3人以上一致					
		係り先のみ		粗い分類		細かい分類		係り先のみ		粗い分類		細かい分類	
		数	率	数	率	数	率	数	率	数	率	数	率
A	44	23	0.52	17	0.39	16	0.36	35	0.80	26	0.59	24	0.55
B	57	29	0.51	21	0.37	17	0.30	46	0.81	40	0.70	37	0.65
C	37	28	0.76	17	0.46	17	0.46	32	0.86	25	0.68	25	0.68
D	55	38	0.69	31	0.56	27	0.49	50	0.91	43	0.78	42	0.76
E	47	26	0.55	12	0.26	9	0.19	37	0.79	26	0.55	23	0.49
F	46	25	0.54	20	0.43	16	0.35	33	0.72	31	0.67	29	0.63
G	27	16	0.59	7	0.26	7	0.26	20	0.74	15	0.56	14	0.52
H	35	21	0.60	17	0.49	14	0.40	29	0.83	24	0.69	22	0.63
合計	348	206	0.59	142	0.41	123	0.35	282	0.81	230	0.66	216	0.62

今回我々は、この中の8件を用いて、論理関係の体系(3.1節)、および詳細化アウトラインに対する論理関係の付与手続き(3.2節)に基づいて、どの程度頑健かつ一貫して論理関係・修辞構造を特定できるかを調査した。

まず、第1、第2著者、他の2名の合計4名(いずれもFY2013データの作成者ではない)が、各自が独立に修辞構造情報を付与した。その後、全員分を横並びに見ながら各自が軽微な誤りを修正した。対象としたサンプルの記述統計、および4名の判断の一致の程度を表3に示す。表中の「粗い分類」は表1および表2における大分類(8種類)を指し、「細かい分類」はそれらの表の右端の具体的な関係の種類(20種類)を指す。

係り先および関係の一致率は、文書ごとに大きく異なった。CとDは細かい分類まで見ても相対的によく一致したが、A, B, E, F, Gでは係り先の特定も難しかった。係り先が3人以上一致した282個の要素のうち、その全員が同じ粗い関係を付与したのは190個(67%)のみであった。新たな不一致を生じた92個のうち、単核関係と多核関係に判断が分かれた要素が24個あり、削除テストの改善の必要性が示唆される。また、単核関係の「D1. 著者の解釈/意見」と「D5. 情報付加」の間の相違が23個あった。これらは、当該要素が著者あるいは他者の意見であるか否かを容易に判断できなかったことに起因する。詳細化アウトラインの作成時に、この点を明確するように教示することも必要である。係り先の一致率は低かったが、修辞構造の根となる要素については、8件中4件について4人が一致、7件について3人以上が一致した。ボトムアップに関係を同定しても主たる意見は概ね特定できた。

粗い関係が3人以上一致した230個の要素のうち201個(87%)について、その全員が細かいレベルでも同じ関係を付与した。このうち28個が「D5. 情報付加」の下位分類に関するものであり、特に「背景」、「例示」、「詳細化」で判断が分かれる例が多かった。決定木における条件の精査が必要であることが分かった。

5 おわりに

我々は、高等教育機関において文章の作成・推敲能力を涵養するための方法論を設計・検証している。本稿では、これまでに作成した、詳細化アウトライン、修辞構造理論に基づく論理関係の体系、および詳細化アウトラインに対する論理関係の付与手続きについて、設計手順と現状の課題を示した。

今後は、教育現場にこれらのツールを導入し、実践を通じてその効果の多寡、改善すべき点を明らかにする[9]。また、高品質な詳細化アウトラインのデータを蓄積し、学習者向けの教材として活用すること、およびこれを利用して修辞構造の自動推定ツールを開発することを計画している。

謝辞: 本研究は科研費基盤研究(B)(課題番号:25282060, 代表:椿本弥生)の支援を受けた。

参考文献

- [1] L. Carlson and D. Marcu. Discourse tagging reference manual. Technical Report 545, Information Sciences Institute, University of Southern California, 2001.
- [2] 中央教育審議会. 学士課程教育の構築に向けて(答申), 2008. http://www.mext.go.jp/b_menu/shingi/toushin.htm.
- [3] L. Faigley and S. P. Witte. Analyzing revision. *College Composition and Communication*, Vol. 32, No. 4, pp. 400-414, 1981.
- [4] 平尾努, 西野正彬, 安田宜仁, 永田昌明. 談話構造に基づく単一文書要約. 言語処理学会第19回年次大会発表論文集, pp. 493-495, 2013.
- [5] 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.
- [6] 加藤翔. 文章作成教育に向けたアウトラインに対する修辞構造情報付与の試み. 平成25年度公立はこだて未来大学卒業論文, 2014.
- [7] W. C. Mann and S. A. Thompson. Rhetorical structure theory: Toward a functional theory of text organization. *Text*, Vol. 8, No. 3, pp. 243-281, 1988.
- [8] テクニカルコミュニケーター協会. 日本語スタイルガイド第2版. テクニカルコミュニケーター協会出版事業部会, 2011.
- [9] 富永敦子, 大塚裕子. 詳細化アウトラインを用いた文章表現授業の実践. 日本教育工学会研究会, 2015. (to appear).