

評論文における対比構造とその自動検出

佐野 正裕 佐藤 理史 松崎 拓也

名古屋大学大学院工学研究科

m_sano@nuee.nagoya-u.ac.jp

1 はじめに

文章を理解するためには、その文章の構造を把握することが不可欠である。近年では、議論の構造の把握を目的とする研究領域 (Argument Mining) において、議論を構成する要素 (主張や理由) が文章中でどこに述べられているかを特定する手法が研究されている [3]。このほかにも、テキスト間の関係を示す言論マップの生成を目的として、Web から「同意」「対立」「対立根拠」など6種類の情報を含む文を抽出する手法 [9] や、Web 情報の俯瞰的把握のために、主要・対比・対立文を抽出する手法 [6] などが提案されている。

文章の理解を問う国語入試問題においても、文章構造の把握は、正解を導くための鍵となる。入試問題でよく出題される評論文においては、著者の主張を展開するための典型的な文章構造が数種類存在する [7]。そのような文章構造の一つに、2つの事柄を対比・比較して違いを際立たせる**対比構造**がある。

我々は、対比構造の自動検出の可能性を探るために、入試問題の評論文を対象に、文章中に含まれる対比構造の特徴を調査した。そして、その結果に基づいて対比構造検出システムを試作した。本論文では、これらの内容について報告する。

2 対比構造と評論読解問題

一般に、対比構造は、2つの事柄を対比・比較して述べる構造で、その違いを際立たせる効果を持つ。たとえば、次の文では、「予言」と「予測」を対比し、それらの性質の違いを述べている。

- (1) 予言とは「超越的」で「非科学的」な根拠や手法により、未来を「確定的」に語るものであるのに対して、予測は「経験的」で「科学的」な根拠と手続きにもとづいて、未来を「可能性」や「蓋然性」において語る。(出典: 弘前大学入試 [2])

本研究では、**対比関係**にある2つのテキスト (**対比要素**) によって構成される構造を**対比構造**とみなす

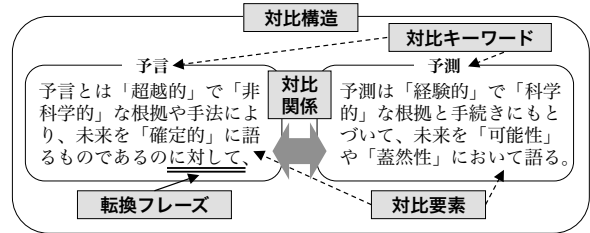


図 1: 文 (1) の対比構造

(図 1)。1つ目の対比要素の末尾、または、2つ目の対比要素の先頭には、要素が切り替わることを明示的に示すフレーズが存在することが多い。これを**転換フレーズ**と呼ぶ。上記の文 (1) では、「に対して」が、転換フレーズである。多くの場合、それぞれの対比要素には、比較対象を端的に表すフレーズが存在する。これを**対比キーワード**と呼ぶ。上記の文 (1) では、「予言」と「予測」が対比キーワードである。

対比構造は、評論読解問題を解くために把握すべき重要な構造の一つとみなされている。たとえば、福嶋 [8] は、文章を読解するために把握すべき関係として、対比関係を含む次の3つの関係をあげている。

- **同等関係** 言いかえる力。抽象と具体の関係。
(例) バナナ ⇔ 果物
- **対比関係** くらべる力。
(例) このボールは大きい ⇔ あのボールは小さい
- **因果関係** たどる力。原因と結果の関係。
(例) 宿題を忘れた ⇔ 叱られた

板野 [4] は、評論文の読解に必要な文の接続パターンを11種類 (順接、逆接、要約・言い換え、並立、添加、比較、譲歩、対比、選択、例示、転換) に整理し、特に対比要素の整理の重要性を強調している。

このように、評論文の読解で対比関係・構造の把握が重要視されるのは、著者の主張を展開するための主要構造として、対比構造がしばしば使われるからである。この使われ方には、対立意見と対比することによって著者の主張を際立たせる**対比型** (文献 [5] の「二項対

(本文)

なぜ日本ではこのような教育方法が行なわれていたのでしょうか。一つには、東洋の特性ということが言えると思います。東洋には「体を整えることによって精神を整える」という伝統が、インド、中国、日本と綿々と続いています。

西洋にも身体と精神を結びつける思想がないわけではないのですが、東洋ほど色濃いものはありません。むしろ西洋においては、知的に高度であるということ、姿勢を整えて礼儀正しくするという事は直結していないと言えます。

(中略)

西洋では、最も大切なのは議論の内容に集中することなので、相手が頼杖をついていようが何をしようが、姿勢で「君、それは自分に対して失礼じゃないか」と言う人はいないのです。しかし、礼節を重んじる東洋においては、姿勢を整えて臨むことが非常に重要なことなのです。これは、東洋と西洋の大きな違いと言えます。

(設問) 東洋と西洋の大きな違いとありますが、筆者は西洋と比べて東洋はどうかと言っていますか。次の文に続くようにし、「身体と精神」という言葉を使って説明しなさい。

西洋では、姿勢のことは気にせず、議論の内容に集中することが大切なのに対して、

(解答例) 東洋では、身体と精神を結びつける思想があるので、姿勢を整えて臨むことが非常に重要である。

図 2: 読解問題の具体例 (富山県公立高校入試 [1])

立) と、2つの意見を対比させ、その折衷案を提案する折衷型(文献 [5] の「弁証法」)がある。

図 2 に、解答に対比構造の把握が必要な読解問題の具体例を示す。この設問に解答するためには、この文章で対比されている「東洋」と「西洋」の違いについて、それぞれの内容が文章中のどこに書かれているのかを見つけることが必要である。

3 対比構造にみられる特徴

評論文の対比構造にどのような特徴がみられるかを明らかにするために、評論文を本文とする入試問題 300 問を観察した。本節では、その結果をまとめる。

3.1 対比キーワード

入試問題の評論文の対比キーワードには、よく用いられるペアが存在する。たとえば、「自然」と「人間」や、「都市」と「田舎」などである。参考書(たとえば [10]) には、このようなペア(頻出対比キーワードペア)のリストが示されている。

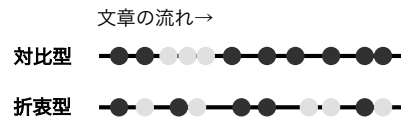


図 3: 期待される対比キーワードの分布

表 1: 転換フレーズ

接続表現	に対して, これに対して, それに対して, 一方, 他方, 逆に, しかし, でも, ところが, だが, が, ではなく, よりも, ならば, だとしたら
副詞	むしろ, 逆に
呼応表現	たしかに～しかし, 一方は～もう一方は～, ～の場合～の場合
間接表現	違い, 対比, 比較, 2つの～, 異なる

対比キーワードのペアは、しばしばその表層構造を共有する。これは、2つのキーワードに表層上の類似性を持たせることにより、対比関係であることを明確にする意図があると考えられる。たとえば、「欧米的な独立的自己観」と「日本的な相互協調的自己観」(出典: 鹿児島県公立高校入試 [1]) は、その構造「～的な～的自己観」を共有する。

3.2 対比キーワードの出現と分布

対比キーワードは、「～は」のような提題助詞でマークされることが多い。併せて、2つの要素を比較することから、「～と～」や「2つの～」のような、並列表現において出現することも多い。

対比キーワードは、2節で述べた対比型と折衷型のそれぞれで、文章中に出現する位置の傾向が異なる(図 3)。

- **対比型** 筆者の主張側の対比キーワード(●)は、文章全体に広く出現するのに対し、対立側の対比キーワード(○)は、文章の一部にしか出現しないことが多い。
- **折衷型** 2つの対比キーワード(●・○)は、対になって出現することが多い。

3.3 転換フレーズ

転換フレーズとして、表 1 に示すような表現が用いられる。その多くは接続詞・接続表現であるが、「違い」のように、間接的に示唆する表現もある。

3.4 対比構造の大きさ

対比構造の範囲は、1つ目の対比要素の開始点から2つ目の対比要素の終了点までである。これを対比構造の大きさとするとき、文章中には、色々な大きさの対比構造が出現する。それらは、次のように大別できる。

- **一文内の対比** 対比構造は一文内に収まる。
- **段落内の対比** 対比構造は一段落内に収まる。
- **複数段落の対比** 対比構造は複数段落にわたる。

転換フレーズの出現位置と対比構造の大きさには、次のような傾向が観察される。

- 転換フレーズが文中に出現する場合
→ 一文内の対比となる
- 転換フレーズが段落の先頭以外の文頭に出現する場合
→ 段落内の対比となる場合が多いが、複数段落の対比となる場合もある
- 転換フレーズが段落の先頭に出現する場合
→ 複数段落の対比となる

一文内の対比の場合、対比構造の大きさは、その一文の大きさと一致するのが普通である。しかしながら、段落内の対比の場合は、段落の一部が対比構造となっている場合があり、対比構造の大きさは、必ずしも段落の単位と一致するとは限らない。複数段落の対比の場合は、転換フレーズが段落の先頭にある場合においても、前後の2段落がそれぞれこのまま対比要素になるとは限らない。すなわち、段落内および複数段落の対比構造の範囲は、転換フレーズの位置と段落単位だけからは同定できない。

4 対比構造の自動検出

3節に示した特徴を利用して、文章中の対比構造を自動検出するシステムを作成した。

4.1 問題設定

ここでは、対比構造の検出問題を、与えられた文章から、表2のような形式の表を作成する問題として設定する。この表は、2つの対比要素それぞれに対して対比キーワードを付与した、4要素からなる表である。ここで、一方の対比キーワードは、所与とする。すなわち、本システムの入力は、評論文の文章と一方の対比キーワードであり、その出力は、表の残りの3要素を埋めた表である。なお、一つの文章に複数の対比構

表 2: 図 2 の本文の対比構造

(a) 本文の前半

西洋	東洋
西洋にも身体と精神を結びつける思想がないわけではないのですが、東洋ほど色濃いものではありません。	東洋には「体を整えることによって精神を整える」という伝統が、インド、中国、日本と綿々と続いています。

(b) 本文の後半

西洋	東洋
西洋では、最も大切なのは議論の内容に集中することなので、相手が頰杖をついていようが何をしていようが、姿勢で「君、それは自分に対して失礼じゃないか」と言う人はいないのです。	(しかし、) 礼節を重んじる東洋においては、姿勢を整えて臨むことが非常に重要なことなのです。

造が存在する場合は、それらすべてを検出対象とする。表2は、図2の文章から検出されるべき2つの対比構造を示している。

本システムでは、表層的な手がかりで実現可能な範囲を実装した。すなわち、対比構造を検出する手がかりとして転換フレーズと段落区切りを利用し、「一文内の対比」と「段落内の対比」を検出することとした。対比要素の同定では、対比構造の大きさと文単位、段落単位が一致することを仮定した。

評論文の文章としては、公立高校入試や大学入試で出題された評論文の本文の32文章を用いた。それぞれの文章に対して、段落情報はあらかじめ付与し、文の認定は句点に基づいて行うものとした。

4.2 検出手順

本システムは、次の手順で、対比構造を検出する。

1. **転換フレーズの検出** 予め用意した転換フレーズのリスト(表1の**太字**)を用い、形態素列パターンマッチングにより、入力された文章に含まれる転換フレーズを検出する。
2. **対比要素の特定** 検出した転換フレーズの出現位置に基づいて、対比構造の範囲を決定する。転換フレーズが文中に出現した場合は、一文内の対比とみなし、転換フレーズの直後で文を分割して文の前半と後半をそれぞれ対比要素とする。転換フレーズが(段落内の)文頭に出現した場合は、段落内の対比とみなし、転換フレーズの直前で段落を分割して段落の前半と後半をそれぞれ対比要素とする。転換フレーズが段落の先頭に出現した場

表 3: 検出結果

	一文内の対比	段落内の対比
転換フレーズの特定	19/25 (76%)	32/35 (91%)
対比要素の特定	19/25 (76%)	16/35 (46%)
対比キーワードの抽出	11/25 (44%)	6/35 (17%)

合は、複数段落の対比とみなし、抽出対象とはしない。

3. **抽出対象の判定** 所与の対比キーワードがどちらかの対比要素に含まれる場合、この対比構造を抽出する。
4. **対比キーワードの抽出** 所与の対比キーワードを含まない対比要素から、もう一方の対比キーワードを次の優先順位で探し、抽出する。
 - (a) 頻出対比キーワードペアのリストにおいて、所与の対比キーワードとペアをなす表現
 - (b) 所与の対比キーワードと類似した構造を持つ表現 (所与の対比キーワードを構成する形態素を変数化したパターンを使用して探す)
 - (c) 助詞「は」でマークされる形態素

4.3 システムの出力結果と考察

入力として与えた 32 文章 (対比キーワード 52 件) に対して、システムは、一文内の対比を 25 個、段落内の対比を 35 個出力した。この出力を観察し、検出手順のステップごとに期待される結果が得られているかどうかを確かめた。その結果を表 3 に示す。

一文内の対比は、転換フレーズの特定誤りがやや多かった (6/25) もの、転換フレーズが特定できた場合は、対比要素も正しく特定できた。これは、一文内の対比は、対比構造の範囲が一文と一致することを意味し、3 節での観察結果と一致する。

段落内の対比は、転換フレーズの特定は多くの場合に成功したが、対比要素の特定では、その半分 (16/32) が失敗した。最終的に対比キーワードまで特定できたものは、一文内の対比では 44%、段落内の対比では 17%であった。

これらの結果から、次の 3 つの課題が浮かび上がる。

第一の課題は、「転換フレーズの特定」に関わる課題である。今回利用した転換フレーズのリストは、対比構造の転換点を示す典型的な表現を集めたものであるが、必ずしも対比構造の転換点となるとは限らない。そのため、転換フレーズの誤認識は避けられない。特

に、文中に現れる転換フレーズの誤認識が多い。この問題に対しては、分割した後の文の前半部と後半部の長さがほぼ等しくなることを条件にするなどのヒューリスティックを併用する必要がある。

第二の課題は、「対比構造の開始点と終了点の決定」に関わる課題である。一文内の対比の場合は、対比構造の大きさと文の単位がほとんどの場合一致するため問題は生じないが、3 節でも述べたように、段落内の対比の場合は、対比構造の大きさと段落の単位が必ずしも一致しない。現在のシステムは、これが一致することを仮定したアルゴリズムとなっているため、対比構造の大きさの同定に失敗する。この問題を解決するためには、文間関係や段落間関係の認識が不可欠と考えられるが、その実現は容易ではない。

第三の課題は、「対比キーワードの抽出」に関わる課題である。現在のシステムは、表層的な手がかりのみを利用するが、表層的な対応がないキーワードペアであっても、反義語のような意味的な対応がとれるキーワードに対しては、意味情報の利用により抽出できると考えられる。そのほか、名詞句以外のキーワードペアのような、キーワードの単位特定が不十分となるキーワードに対しては、文節構造や出現回数を利用するといった手法も検討する必要がある。

謝辞 本研究は JSPS 科学研究費基盤研究 (B) 「日本語文章の構造モデルとその段階的詳細化による文章自動生成機構」 (課題番号 18H03285) の助成を受けている。

参考文献

- [1] 2018 年受験用 全国高校入試問題正解 国語. 旺文社, 2017.
- [2] 現代文問題データベース Vol.6. 明治書院, 2017.
- [3] Eckle-Kohler Judith, Roland Kluge, and Iryna Gurevych. On the role of discourse markers for discriminating claims and premises in argumentative discourse. *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pp. 2236–2242, 2015.
- [4] 板野博行. 現代文ゴロゴ 解法公式集 (2) 国公立・私立大編. スタディーカンパニー, 2017.
- [5] 今井健仁. 現代文の解法 第 3 版 (東京大学への道). データハウス, 2015.
- [6] 河原大輔, 乾健太郎, 黒橋禎夫. Web 情報の俯瞰的把握のための主要・対比・対立文の抽出と集約. 言語処理学会第 16 回年次大会発表論文集, pp. 134–137, 2010.
- [7] 神田邦彦. 現代文 標準問題精講. 旺文社, 2015.
- [8] 福嶋隆史. 高校受験 [必携] ハンドブック 国語読解 [完全攻略]22 の鉄則. 大和出版, 2014.
- [9] 水野淳太, Eric Nichols, 渡邊陽太郎, 村上浩司, 松吉俊, 大木環美, 乾健太郎, 松本裕治. 言語マップ生成技術の現状と課題. 言語処理学会第 17 回年次大会講演論文集, pp. 49–52, 2011.
- [10] 村上翔平. 高校 現代文をひとつひとつわかりやすく. 学研プラス, 2017.