

論説文における論証構造解析と視覚化インターフェースの提案

川本 大輔† 田村 直良††

† 横浜国立大学大学院 環境情報学府
†† 横浜国立大学大学院 環境情報研究院
{kawamoto,tam}@tamlab.ynu.ac.jp

1 はじめに

本研究では、論説文における話題の移り変わりに着目した筆者の論証の進め方についてのモデル化し、それに基づいた文章解析を行う。また、応用として解析結果を視覚化して提示し、人間が文章評価を行う際の支援となるようなユーザインターフェースを提案し実現する。

文章とは、書き手の思考を読み手に伝達するための手段であり、筆者の主張を説得力を持たせ読み手に伝えることを目的とした文章が新聞社説をはじめとした論説文章である。

この論説文章の性質を記述する手段として修辞構造理論 [3] などの文章構造解析手法がある。修辞構造理論に基づいた文章構造解析の研究としては小野ら [2] は日本語の論説文を対象に、接続詞に着目して文脈表現を形式化している。福本 [1] は論説文を筆者の主張という観点から捉えて文章の構造化を行った。また、春日 [5] は修辞構造解析と論旨レベルの解析を二段階的に組み合わせた解析を行っている。これまでの修辞構造理論に基づいた解析手法には次のような問題点がある。

- 隣接した文同士でスパンを形成してまとめていく構造のため、隣接しない文同士の詳細な関係が分からない
- 修辞構造木の根の付近では、修辞部分木同士を修辞関係で結び付けることが困難である
- 論証の展開の構造、局所的な論旨の移り変わりまでは把握できない

これまでの文章解析は修辞構造理論に基づいての木構造解析が多かったが、上記の問題を考えると、単純な木構造ではなく筆者の論旨の展開も表現できるような文章構造が必要である。

そこで本研究では、論説文における筆者の論証の展開に着目し、一つの文をノードとし、その関係をアークとしたネットワーク構造で表現する構造を提案する。

また、修辞構造理論は記述理論としては有力であるが、形式的な定義や定式化については言及されて

いないため、その応用について述べられているものは少ない。本研究ではこの点に着目し、木構造ではない文章構造を提案し、解析結果の応用として文章評価支援のためのグラフィカルユーザインターフェースを提案する。

2 文章構造の特性とモデル化

2.1 主題の結束構造

主題 (theme) とは一文において話題の中心となる語句であり、日本語では主に係助詞「は」などによって提示される。また、文中の主題以外の部分を題述 (rheme) と呼ぶ。新たに主題として話題を提示することにより読者の注意を引き付け、後に再び同じ主題を用いたり、題述として登用した情報を後で主題として用いることで注目する話題として強調される。ここで、新しく話題として表出する情報を新情報 (new information)、既に言及された既知の情報を旧情報 (old information) と呼ぶ。実際の論説文章では、この新情報と旧情報の主題・題述の連鎖構造により結束性を強めながら筆者の最終的な主張が読者にスムーズに受け入れられるような構造を形成している。

2.2 接続表現

文と文をつなぐ接続表現は文章において筆者の論旨の展開を最も明示的に表現する手段である。用いられる接続表現により文間の接続関係を同定することができる。ここでは、表 1 のように接続表現を分類する。ただし、接続表現が現れる文を調査したところ全体の文の約 13% ほどしかなく、これだけで文章構造を解析するには不十分であると考えられる。

2.3 文のムードタイプ

文のムードタイプとは文末の表現に現れるその文の性質を記述したものであり、ここでは、客観的に事実を述べる叙述文と筆者の判断が含まれる判断

接続関係	接続表現の例
理由	その理由は、なぜなら、というのは
例示	たとえば、今回は
換言	いわば、つまり、すなわち
強調	言うまでもなく、まして、なおさら とくに、さらに、とりわけ もちろん、ただ
累加	しかも、それに、おまけに なお、もっとも、その上
説明	それは
逆接	しかし、残念ながら、だが でも、ところが、けれども
並列	また、同時に
選択	もしくは、あるいは、または
対比	いっぽう、これに対し(て)
提起	問題は、問題として
因果	その結果、そのため、このため これでは
結論	したがって、だから、結局(は)
順接	そして、そこで
相反	それでも、それなのに、それより むしろ
一般化	このように
根拠	だからこそ、これこそ、それこそ
予測	そうすれば、そうなれば
条件	もし、とすれば、だとすれば、ならば そうなら(ば)
展開	ところで、さて

表 1: 接続表現の分類

文、筆者の意見を述べる意見文の3タイプに分類する。このムードタイプによる文間の接続も文間の接続関係を同定する際の指標となると考える。例えば前文が判断文で後文が意見文となる場合は、その文間の関係は“根拠”であると便宜的に定めることができる。このムードタイプの接続による関係を定めたものを表2に示す。

2.4 文章構造モデル化

以上に述べた主題構造、接続表現、ムードタイプによる文同士の関係を複合的に解析するモデルを提案する。このモデルは従来の修辞構造解析のような部分木として結び付ける2分木構造ではなく一文

ムードタイプの順序	文間の関係
叙述 → 叙述	順接
叙述 → 断定	説明
叙述 → 意見	前提
断定 → 叙述	補足
断定 → 断定	順接
断定 → 意見	根拠
意見 → 叙述	補足
意見 → 断定	補足
意見 → 意見	順接

表 2: ムードタイプの順序による文間関係の同定

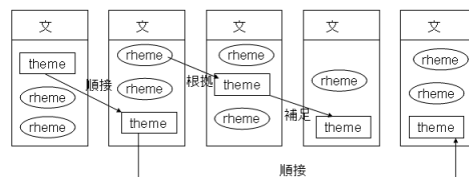


図 1: 文章構造の表現

をノードとして関係によりリンクが張られたネットワーク的な構造で表現する。この図を図1に示す。

3 文章構造解析

3.1 表層からの情報抽出

まず、文章に対して形態素解析と構文解析を行い品詞情報を抽出する。次に表層の表現から、関係解析に必要な意味的な情報を抽出する。これには以下のものがある。

- 主題 (theme) の抽出

ある文における主題を決定する方法として、助詞、初出現の「八格」、存在しない場合は「ガ格」の直前に出現する名詞または名詞句を抽出する。この際、主題の連鎖構造を明確に検出できるように「もの」、「こと」など形式的な語が主題候補として抽出された場合は、これを除外する。また、注目する文中に主題候補が存在しない場合は前文の主題が継続しているものとする。

- 題述 (rheme) の抽出

上記の操作により主題として検出されなかった、名詞、名詞句の集合を題述として抽出する。

- 接続表現の抽出

文頭に存在する接続詞、または接続表現を抽出する。

- ムード情報の抽出

文末の表現から叙述文、意見文、断定文を判定し文のムードタイプ情報を抽出する。

3.2 主題連鎖構造解析

文章の中のある文の主題がそれより前の文のどの部分（主題・題述）から引用されているかという主題の連鎖構造を解析する。具体的なアルゴリズムは、まず直前の文の主題・題述を参照し、語彙連鎖が認められる場合それら2文の間にリンクを形成する。直前の文に語彙連鎖がない場合はさらに一つ前の文を参照し、先頭の文まで順に遡ってリンク先を決定する。

さらに、リンクの張られた文間の関係同定も重要である。文同士の関係はムードタイプを基に前章で述べた接続関係で同定する。

3.3 接続表現による関係解析

主題構造による関係同定に加えて、隣接する2文間においてはその接続表現による関係も同時に同定を行う。同定する関係名は前章で述べた接続表現の分類を基に関係付けを行う。

4 文章構造解析ユーザインターフェイス

4.1 文章評価の観点から

文章構造解析の結果をアプリケーションとして応用することを考えたときに論説文の文章評価という観点で応用が期待できるのではないかと考える。この文章評価について考察した研究として渡部ら [7] の研究があり、小論文の評価を行う際の評価基準を分類し解析を行っている。石岡ら [6] はこの分類された観点到に着目し、「修辞」と「論理構成」、「内容」の大きく3つに分類し、それぞれに関して、小論文の自動採点という形で文章評価を実現している。

このうち、本研究による文章構造の解析結果を「論理構成」という観点から文章評価へと応用できる。石岡らの研究では、この「論理構成」に関する評価の基準としては、接続詞の出現パターンが新聞

社説における出現パターンと類似しているかを評価基準としている。

この「論理構成」に着目した評価基準として、接続表現の出現パターンだけではなく、本研究で示した論証構造の解析結果を応用することができる。実際の文章に対して点数をつける採点については対象外とし、評価の自動化、または評価者の負担軽減に貢献できるようなインターフェイスの作成を目的とした。

4.2 評価観点の考察と機能の実装

ここでは、文章評価の「論理構成」という観点をさらに掘り下げて考察し、必要な機能として実装する方法について述べる。

論説文や小論文などの文章において中核をなすものは、その文章で何を述べたいかという筆者の主張である。この観点は、福澤 [8] を参考にすると、議論をする際に主張を根拠が支える構造になっているかという考え方である。この観点を実現する機能としては、ムードタイプの分類による文の割合を提示する。

文章における、主題構造の結束性も論証の構造を評価する重要な観点である。Miltakakiら [4] によると、主題として現れる局所的な話題の結束性が文章評価に寄与するとして、Rough-Shiftと呼ばれる主題構造に結束性のない文章はその評価が低くなることを示している。これを提示する機能として主題・題述構造の連鎖の個数が文章中にどれくらい含まれているかを表示する機能を実装する。また、接続表現による文間のつながりも提示する機能を実装する。

4.3 ユーザインターフェイスとしての実装

前節までに述べた項目を実現し、視覚的に提示するグラフィカルユーザインターフェイスをMicrosoft Visual C++ 6.0を用いて構築した。さらに評価支援ツールとしての有用性を高めるためにインターフェイスには入力文章を表示し、文のムードタイプごとに色を分けて直感的に構造が理解しやすいようにした。また、操作として、文を選択するとどのように主題・題述連鎖や接続表現による論証構造をたどっているかのリンクを文間の関係ごとに色分けし、ハイライトして表示する機能を実装した。実装したユーザインターフェイスの出力の結果を図2に示す。

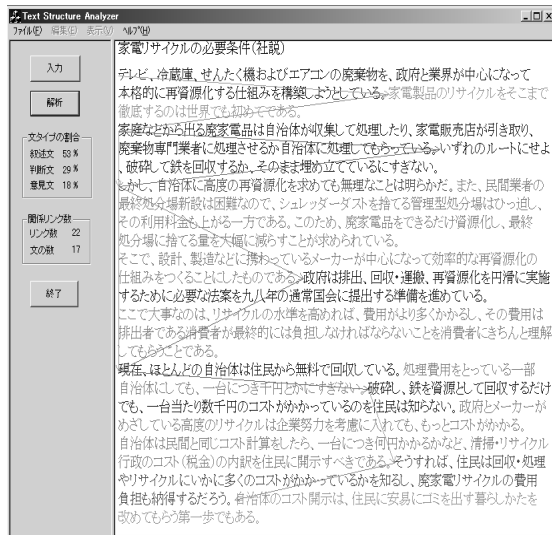


図 2: ユーザーインターフェイスの出力結果

5 解析結果の評価と考察

本章では3章で述べたアルゴリズムに基づいて、実際に新聞社説10記事¹について解析を行い、その結果を評価・考察する。

5.1 主題の抽出

各文について人手で主題となる語句を選定し、今回の手法で抽出した結果の正解率を集計し73%の正解率を得た。主題が係助詞「は」により明示的に提示されている文や動作の主語が主題となるような文においては、概ね正確に主題を抽出することができた。しかし、目的語が主題としての機能を果たす文などでは、今回の手法では抽出できなかった。また、文が重文構造になっていて主題候補が複数存在するような場合においては、初出現のものを主題としたが、必ずしもそうなるとは限らなかった。抽出ルールをさらに細分化することで精度の向上が期待できると考えられる。

5.2 文章構造解析結果

文章構造解析結果についても評価・考察を行う。人間による解析との比較のような定量的な判定は避け、計算機による解析結果を人間がチェックして、明らかにおかしいと判断できるところについて評価したところ全体で72%の正解率となった。接続表現による文間の接続関係はほぼ誤りなく解析できていた。一方、主題連鎖構造による文間の関係解析では、文同士が隣接または近隣に存在する場合には高

い精度で解析が行えていた。しかし、文の位置が離れているときには、正しく関係付けが行えた場合もあったが、関係誤りとして抽出された場合が多かった。文間の距離や文のムードタイプさらに考慮した解析により精度の向上が見込まれる。そもそも主題抽出の段階で誤りがあり、文間の主題連鎖関係が正確に把握できなかったことも課題である。

6 おわりに

本研究では論説文の形式的な構造をふまえた上で、筆者の主張を導き出すまでの論証の展開をモデル化した。その際、筆者の論旨の中心となる話題の変遷に着目し構造化を行った。また、解析の応用として文章評価の際の支援となるようなインターフェイスを考案し実装した。

参考文献

- [1] 福本淳一, 安原宏. 文の接続関係解析に基づく文章構造解析. 情報処理学会自然言語処理研究会, Vol. 88, No. 2, 1992.
- [2] 小野顕司, 浮田輝彦, 天野真家. 文脈構造の解析. 情報処理学会研究報告, Vol. 70, No. 2, Jan. 1989.
- [3] W. C. Mann. Rhetorical structure theory : Description and construction of text structure. In G. Kempen, editor, *Natural Language Generation*, pp. 279–300. Martinus Nijhoff Publishers, 1987.
- [4] E. Miltsakaki, K. Kukich. Evaluation of text coherence for electronic essay scoring systems. *Journal of Natural Language Engineering* Vol.10(1), 2004.
- [5] 春日隆緒. 文章のセグメント間における関係解析に関する研究. Master's thesis, 横浜国立大学大学院, 2003.
- [6] 石岡恒憲, 亀田雅之. コンピュータによる日本語小論文の自動採点システム. 電子情報通信学会, 信学技報 Vol.102, No.491, 43-48, 2002.
- [7] 渡部洋, 平由実子, 井上俊哉. 小論文評価データの解析. 東京大学教育学部紀要, 第28巻, 143-164, 1988.
- [8] 福澤一吉. 議論のレッスン. NHK出版, 2002.

¹日本経済新聞 1998年の社説