

表層表現に基づく文章構造解析による 論文改訂支援システムの試作

2M-4

中島 靖, 渡川 智浩, 中村 順一, 吉田 将

九州工業大学 情報工学部

1. はじめに 科学技術論文などの論理的文章では、その論旨展開が明確になっており、事実、主張、根拠などがバランスよく述べられていなければならない。より良い論文を作成するためには、よく推敲し、論文を何度も改訂する必要がある。本稿では、このような論文の改訂作業を支援するシステムについて述べる。本システムでは、改訂対象の文章の構造を解析し、その結果を筆者に提示することにより支援を行う。文章構造の解析は、文タイプ(例: 事実、推定)、接続詞(例: よって、しかし)などの表層表現をキーとして、chart パーサにより行う。筆者への提示のため、NEmacs 上でユーザインタフェースを作成した。

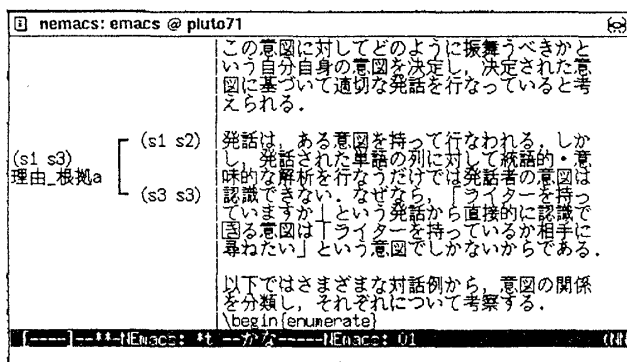


図 1: 文章構造と論文改訂支援の関係

2. 改訂支援方法 論文改訂の支援は、筆者に文章構造を提示することにより行う。例えば、図1に示すように、文間にどのような論理的関係があるか、そして、それらがどのようにまとまって段落となっているかを表示する。図1の例であれば、文 s1, s2 と s3 との間に理由(根拠)の関係があることが示される。論文の筆者は、この表示を見て、システムが判断した論理的関係が意図しているものと一致しているか、また、全体のバランスは適切であるかを判断し、もし問題があれば、記述したい内容にあわせて変更を行えば

⁰Technical Writing Support System based on Text Structure Analysis using Clue Expressions
Yasushi NAKASHIMA, Tomohiro ORIKAWA,
Jun-ichi NAKAMURA, Sho YOSHIDA
Department of Artificial Intelligence
Kyushu Institute of Technology

```
書式: s(文番号, 文タイプ, 接続表現タイプ, 後続文規定表現タイプ).
抽出結果: s( 1, 事実, null, null ).
           s( 2, 可能, 接逆接, null ).
           s( 3, 断定, 接理由a, null ).
           s( 3, 断定, 表理由a, null ).
           #pos=(259,430,396,547)
```

図 2: 文特徴の抽出結果の例

よい。

3. 文章構造の解析方法 文章構造の解析は、(1) 文特徴の抽出、(2) 文章構造の解析の 2 つのステップで行う [中島 92]。これらは処理速度の観点から C 言語によりインプリメントした。

3.1 文特徴の抽出 構造解析の対象の文章から文を切り出し、各文からパターンマッチングにより以下の 3 つの特徴を抽出する [渡辺 86] [中島 92]。

文タイプ: 「である」、「たい」等の文末表現により分類する。各文が表現している大まかな意味を代表する。事実、断定、可能、推測、希望、要求、意見、および、文章構造を規定する L^AT_EX コマンド (itemize, description など) がある。

接続表現タイプ: 「よって」、「しかし」、「ここで」等の接続詞により分類する。文間の論理的な関係を求めるためのキー。理由、逆接、前提などがある。

後続文規程表現タイプ: 「～が n ある」、「以下のものがある」等、この文の後に列挙表現が存在することを明示する表現により分類する。具体化の関係などを判定する手掛りに用いる。

図1の文章に対して文特徴の抽出結果を行った結果を図2に示す。ここで、文 s3 には「なぜなら」、「からである」と2つの接続表現があるので、2つの接続表現タイプを出力している。また、接続表現タイプと後続文規程表現タイプは、すべての文にあるとは限らないので、該当する特徴がない場合は null とした。また、インタフェースにおける構造表示の際に利用する各文の位置情報も抽出した。

```

関係(A, B, 逆接) --> 逆接(A, B, _).
理由_根拠 a(A, B, C) -->
    関係(A, B, C), 文(可能, 理由表現b, null).
逆接(A, B, C) -->
    文(A, B, C), 文(_, 接_逆接, _).
文(A, 理由表現 a, C) --> 文(A, 接_理由 a, C).
文(A, 理由表現 a, C) --> 文(A, 表_理由 a, C).

```

図3: 文章構造の解析用の書き換え規則の一部

```

書式:
(構造の範囲リスト文章構造リスト構造中の各文の位置リスト)
例:
((s1 s3)
(理由_根拠 a<事実・null・逆接>(関係<事実・null・逆接>
(逆接<事実・null・null>(文<事実・null・null>s1)
(文<可能・接_逆接・null>s2))) (文<断定・理由表現 a・null>
(文<断定・接_理由 a・null>s3))) (259 430 396 547))

```

図4: 文章構造の解析結果

3.2 文章構造の解析 文特徴の結果を基に構造解析を行い、文章構造を求める。文章構造は、各文の文特徴を終端記号とする拡張文脈自由文法により規定し、木構造により表現した。構造解析には、ボトムアップ型の chart パーサを用い、書き換え規則は図3に示すように DCG 的な形式で記述した。なお、各非終端記号の引数のうち、アルファベットの大文字が変数であるが、値としてはシンボルのみが許される。図3は、文 a と文 b が逆接の関係があり、これらと次文 c とが理由(根拠)の関係であるような文章(図1)を解析するための書き換え規則である。

最終的な文章構造の解析結果を図4に示す。この構造は、図1に示す文章の文章構造で、4つの複数解のうちの一つである。

4. ユーザインタフェース 本システムのユーザインタフェースは、NEmacs 上に emacs lisp により作成した(図1)。ユーザインタフェースは、利用者が指定したリージョン(段落単位)を前節で述べた文章構造の解析プログラムに送り、その結果を受け取って表示する。

インタフェースは以下の方針により作成した。

1. 文章の改訂の責任はすべて筆者にある。システムはあくまで筆者に改訂のためのヒントを与えるだけとする。そこで、インタフェースは文章構造の解析結果をわかりやすく表示するだけ

とし(図1の左側のウィンド)、実際の文章の変更は、通常の NEmacs の機能により筆者が行う(図1の右側ウィンド)。

2. 文章構造をわかりやすくするため、トップダウン的に表示する。初期画面では、段落全体の解析結果である木構造のルートの直下だけを表示する。ユーザの指示に従って、より下の構造も表示することとした。
3. 文章構造の解析では、一般の自然言語処理システムと同様、(a) 解析結果が複数になる(曖昧さがある)、および (b) 解析結果が得られない、という場合がある。この問題に対処する。

3の(a)については、解を1つずつ表示することにより対処した。方針の2で述べたように、解析木のルートから順に表示することとしたので、ユーザは、解析木の局所的な曖昧さのみを選択すればよい。このため、表示の際の組み合わせの爆発の心配は少ない。

3の(b)については、次のように chart パーサを拡張することにより対処した。解析結果が得られなかった場合には、どのような非終端記号ともマッチする特殊な記号を右辺にもつ規則により、解析結果の chart をスキャンする。この場合、最初に右辺を2個とし、それでも解析できない場合は3個として解析を行う。これにより、段落を2~3個のブロックにまとめ、表示する。なお、実際の論文を調査した結果、段落は6~10文程度であることから、4個以上のブロックなる場合はほとんど構造が解析できなかったと考え、その旨を表示することとした。

5. おわりに 本稿では、文章の構造解析を行い、その結果を表示することにより論文の改訂を支援するシステムについて述べた。現在、システムの支援効果の評価を行っている。また、今後の課題としては、段落単位ではなく、たとえば、節単位で構造を解析すること、文章の変更方法のヒントを表示できるようにすることがある。

参考文献

- [渡辺 86] 渡辺日出雄: 新聞社説を対象とした文章の持つ接続構造の解析, 京都大学大学院工学研究科修士論文(1986).
- [中島 92] 中島靖, 渡川智浩, 中村順一, 吉田将: 表層表現の解析に基づく論文改訂支援方法について, 電子情報通信学会言語理解とコミュニケーション研究会, Vol.92, No.395, pp.25-32(1992).