

表層表現に基づく文章構造解析を利用した  
論文改訂支援システムの試作と評価

甲斐 郷子 中村 順一  
九州工業大学 情報工学部  
〒820 飯塚市川津 680-4

吉田 將  
九州芸術工科大学  
〒815 福岡市南区塩原 4-9-1

科学技術論文などの論理的文章にとって、論旨展開が明確であること、事実、主張、根拠などがバランスよく述べられていることは重要である。しかし初心者にとってこれらの適切さを認識することは容易ではない。本稿では、このような論文作成の初心者に論文改訂作業を支援するシステムの試作と改訂実験によるシステムの評価について述べる。本プロトタイプシステムでは、表層的な情報を基に改訂対象の文章構造を解析し、その結果を筆者に提示することにより支援を行う。改訂実験により明らかになった本プロトタイプシステムの有効性と限界についても論じる。

Technical Writing Support System based on  
Text Structure Analysis using Clue Expressions  
—Prototyping and Evaluation—

Kyoko KAI, Jun-ichi NAKAMURA,  
Department of Artificial Intelligence  
Kyushu Institute of Technology  
680-4 Kawazu, Iizuka, Fukuoka 820, JAPAN

Sho YOSHIDA  
Kyushu Institute of Design  
4-9-1 Siobaru, Minami-ku, Fukuoka,  
Fukuoka 815, JAPAN

A technical writing support system is proposed as an application of text structure analysis technique. For writing technical papers, we should appropriately list the points of its argument, and should balance the number of sentences expressing facts, claims, reasons, and so on. From this point of view, technical writing can be supported by displaying the result of text structure analysis to the author; he can verify appropriateness of his text structure. In this paper, we discuss a prototype of a technical writing support system. This system adopts text structure analysis technique using clue expressions. The result of evaluation by experiments of revising documents is also discussed.

## 1 はじめに

科学技術論文などの論理的な文章を作成する際には、書き手の意図する論理展開が正確かつ効果的に伝わるよう気を配る必要がある。そのため、一旦書き終えた文章を読み直し、論理展開が的確かどうかチェックし、悪ければ書き直すといった改訂作業を何度も行なわなければならない[1]。

しかし、書き手は内容を熟知しているため、論理展開に必要な情報が省略されていても補完して読んでしまいがち、 unnecessary 事柄が記載されていても多少の関連があれば見逃し易い。特に、論文作成の初心者にとっては、意図する論理展開そのものが稚拙であったり、読み手がどのように文章を理解するかについての知識がないために、具合の悪い部分を発見することが大変困難である。

このような文章の改訂作業を支援するシステムとして、現在『推敲』[2]や『日本語 Spell Viser』[3]などが市販されている。これらのシステムは、受身表現や指示詞などの曖昧で読み手が誤解しやすい語句、あるいは誤字脱字や紛らわしい同音語など誤り表記である可能性の高い語句を検出することにより、文章改訂支援を行なうことができる。しかし、これらのシステムが支援するのは語句レベルの改訂であり、論理展開を直接担う文章レベルでの改訂支援システムは未だ開発されていない。

我々は、表層的な表現を基にした文章構造解析の手法[4][5][6]を用いて、書き手に文間接続関係を示す文章構造を提示することにより改訂支援を行なうシステムの構築を行なっている[7][8]。本稿では、論文改訂支援システムの概要と、2つの卒業論文からとった2つの章を対象に行なった改訂実験によるシステムの評価について述べる。

## 2 科学技術論文の改訂作業

論文作成の初心者にとって、自分の書いた文章が論文論旨を正確かつ効果的に伝えているかどうかを判断することは容易でない。そのため論文作成の熟練者が、初心者の作成した論文を直接改訂したり、どのように書き直すべきかを指示することが多い。

熟練者がどのような点に着目して改訂作業を行なうかを明らかにするために、3人の学生が提出した3つの版(計9版)の卒業論文未完成原稿に対して指導教官が行なった改訂作業を調査した。表1に調査結果を示す。ここで、記号や語句に関する改訂とは記号や語句の追加・削除・変

表 1: 教官による改訂

改訂の種類	第1版	第2版	第3版	全体
記号に関する改訂	5	16	1	22
	10.4	7.7	3.3	7.7
語句に関する改訂	20	100	19	139
	41.7	47.8	63.3	48.4
文に関する改訂	7	40	8	55
	14.6	19.1	26.7	19.2
文間接続関係に関する改訂	7	39	2	48
	14.6	18.7	6.7	16.7
スタイル変更	0	4	0	4
	0.0	1.9	0.0	1.4
段落間関係に関する改訂	9	10	0	19
	18.7	4.8	0.0	6.6
全体	48	209	30	287
改訂指示, コメント	41	46	4	91

更を、文に関する改訂とは語順の変更および文の追加・削除・分割・結合・全面書き換えを、文間接続関係に関する改訂とは接続表現の追加・削除・変更や文の順序換えおよび段落の分割・結合を、スタイル変更とは本文・脚注・箇条書間のスタイル変更を、段落間関係に関する改訂とは段落の追加・削除・順序換えおよび章・節の全面書換・構成変更を示している。改訂指示、コメントは、指導教官が具体的な改訂作業は行なわず、改訂を指示したりコメントを述べたものである(例:「参考文献をあげること」「よく分らない?」)。また、表中の上段が改訂(指定)箇所数、下段が全体に対する割合(%)である。文書の大きさは平均で、第一版 27 頁、第二版 39 頁、第三版 46 頁(すべて A4, 12 ポイント)であり、すべてが情報処理ソフトウェア分野の卒業論文である。

表 1のうち語句に関する改訂では、文間接続関係に起因すると思われる文頭・文末表現の変更・追加が 27 箇所あり(例:「を与える。」→「が与えられたとする。」)、文に関する改訂では 27 箇所が接続表現を含んでいた(例:「例えば以下のように表記する。」という文の追加)。これらを含めると、文間接続関係に関する改訂は 102 箇所 35.5 % となる。また、表 1からは早期の版における文間接続関係に関する改訂の比率が高いことが分かる。

もちろん、改訂作業は元の文章に依存するため、この表が一般的な改訂作業の傾向を表わしているとは言えない。しかし、一般的に文間の接続関係と論理展開の間

には密接な関係があることや、特に初心者にとって論理展開を追うことの困難さを考えると、文間接続関係に関する改訂支援は文章全体の改訂支援にとってかなり有意義であると言える。

### 3 文章構造の解析

#### 3.1 文章構造解析の方法

本論文改訂支援システムは、表層的な表現を基にした文章構造解析の手法 [4][5][6] を科学技術論文に適用したものである。この手法は、基本的に深い意味処理は行わず文末表現や接続詞などの表層的な手がかりのみを用いて、文間、あるいは部分構造間の接続関係を決定することにより文章構造を解析する。これは、現状の自然言語処理技術の観点から見て、ロバストな処理を行なうことを前提としたからである。

本システムでは、パターンマッチングによる各文からの特徴抽出と、ボトムアップ型の chart パーサ [9] を用いた文章構造解析の 2 つのステップで文章構造解析を行なう。その際、最近多くの研究者の間で利用されているテキストフォーマッタである L<sup>A</sup>T<sub>E</sub>X の書式コマンドも文章構造解析に利用する。

#### 3.2 各文からの特徴抽出

実用的な論文改訂支援システムに組み込むことを目的としているため、入力文章には事前に一切の手を加えないこととする。そのため、各文から抽出する特徴を抽出が容易な文タイプ、接続表現タイプ、後続文規定表現タイプの 3 つに限定した。以下に、これら 3 つのタイプの文特徴について説明する。

**文タイプ:** 文が表現している大まかな意味を代表するもので、「である」、「たい」等の文末表現により分類する。事実、断定、推測、可能、希望、要求、意見、`tex_begin_*`、`tex_end_*`、`tex_item` の 10 個のタイプを用意した。

**接続表現タイプ:** 文間の論理的関係を求めるための手掛りとなるもので、「よって」、「しかし」等の接続詞や文頭の副詞、「～ためである」等の接続表現により分類する。大きく、理由、逆接、前提、添加、詳細化、具体化、主題/題述、並列、言い換えの 9 つのタイプを用意した。

**後続文規定表現タイプ:** 「～が  $n$  ある」、「以下のものがある」、「～を考えてみよう」など、この表現を持つ文の後に列挙表現または推測文列が存在することを明示するもので、具体化の関係を判定する手がかりとなる。列挙表現が後続するタイプ 1 と推測文列が後続するタイプ 2 を用意した。

#### 3.3 文章構造解析

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 (以下ルールと呼ぶ) により文間あるいは部分構造間の接続関係として規定し、木構造で表現することとした。接続関係としては、前述した接続表現タイプに加え、推測文列、項目規定文部、`TeX_図表`、`TeX_列挙部`、`item` 部を用いる。

文章構造の解析には、ボトムアップ型 chart パーサを用いる。これは、文章構造解析に失敗した場合でも、既に解析されている部分構造を支援システムで利用するためである。自然言語処理において解析に失敗した場合、すなわち全体が 1 つの構造にまとまらなかった場合については、それ以上の処理は行えないのが通常である。しかし、実用的な改訂支援を行なうという立場からはどんな文章でも解析できなければならないし、また 1 つの構造にまとまった文章が良い文章とは限らない。そのため、解析に失敗した場合、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を用いて、解析結果の chart を再スキャンし、段落を 2～3 個のブロックにまとめる<sup>1</sup>。

## 4 論文改訂支援システム

#### 4.1 システム構成

論文改訂支援システムの構成を図 1 に示す。本システムは以下の 3 つのモジュールから成る。

**文特徴抽出モジュール:** ユーザインタフェースから受け取った文章 (1 つの形式段落) を 1 文ごとに分割した後、パターンマッチングを用いて各文の持つ表層的特徴を抽出する。

**文章構造解析モジュール:** 文特徴抽出モジュールにより抽

<sup>1</sup>実際の論文を調査した結果、1 段落の平均文数は 6～10 文程度であったことから、4 個以上のブロックとなる場合はほとんど構造が解析できなかったと考え、ほとんどの無い文章構造を表示することは行わず、その旨を表示するだけとした。



表 2: 各文からの特徴抽出結果

表層的特徴	タイプ	個数	タイプ	個数
文タイプ	断定	22	推測	2
	可能	1	希望	0
	要求	0	意見	2
	tex_*	33	事実	114
接続表現タイプ	理由	3	逆接	1
	前提	14	詳細化	0
	具体化	14	添加	2
	主題/題述	14	並列	8
	言い換え	1	その他	12
後続文規定表現タイプ	タイプ 1	20	タイプ 2	0

表 3: 接続関係抽出結果

文間接続関係名	個数	文間接続関係名	個数
理由	8	逆接	1
前提	22	添加	0
詳細化	14	具体化	24
主題/題述	4	並列	11
言い換え	2	推測文列	0
項目規定文部	2	TeX_図表	0
TeX_列挙部	2	item 部	0
(1文1段落)	7		
何か関係が必要	17	まとまりがない	6

## 5 システムの評価

本システムの評価を行なうため、卒業論文2つから1章ずつとりだし(2章, 46段落, 174文), 改訂実験を行なった。被験者として、手作業で改訂を行なうグループに教官2名と学生4名を、本システムを使用するグループに学生4名を配置した。1人あたりの改訂所要時間は、教官が4.5時間, 手作業者が3時間, システム使用者は2.5時間であった。

まず, 対象とした文章を本システムで処理した結果を表 2, 3に示す。表 2は文タイプ, 接続表現, 後続文規定表現の, 表 3は文間接続関係の抽出結果を示している。このうち目試により他のタイプが望ましいと判断したものは, 文タイプで6, 接続表現タイプで25, 後続文規定表現タイプ6ほどあった。

次に, 1段落1人当たりの改訂数を表 4に示す。この表

表 4: 1段落1人当たりの改訂数

トップノードの表示	構造の曖昧性	段落数	教官	手作業	システム使用
(a) 段落にまとまりがない		6	5.8	0.2	2.6
(b) 何か関係が必要	1	14	2.0	0.2	0.9
	2	3	3.5	0.2	1.4
	5	1	3.5	2.8	0.5
(c) 段落にまとまりがある	1	19	1.9	0.3	0.4
	2	1	2.0	0.0	0.6
	4	1	7.0	0.3	3.1
	8	1	2.0	2.5	2.5

は, システムが提示する全体の文章構造をトップノードの表示と文章構造の曖昧性で分類し, その分類毎に1段落1人当たりの改訂数を表したものである。1段落1人当たりの平均改訂数は教官が2.7, 手作業者が0.3, システム使用者が1.0であった。教官は, (c) 段落にまとまりがあり構造の曖昧性が4の時および(a) 段落にまとまりがない時に多くの改訂をしており, システム使用者も同様である。しかし, 手作業者の場合は(c)で曖昧性が8と(b)での曖昧性が5の場合に多く改訂している。このことから, 本システムを使うことにより, 熟練者と同様の改訂箇所に着目できると言える。

最後に, 1人当たりの改訂数を表 5に示す。この表の見方は表 1と同じである。手作業者に比べシステム使用者は3倍強の改訂を行なっている。これは, システムを使うことにより, 容易に改訂箇所を発見することができたためであると思われる。しかし, 具体的な改訂内容を見ると, 語句や接続表現を付加・変更・削除する小規模な改訂で終わっており, 大規模なより良い改訂に結びついていない場合が多かった。

## 6 結論

本稿では表層表現を基に文章構造解析を行なった結果を提示することにより, 科学技術論文の改訂を支援するシステムについて述べた。改訂実験による本システムの評価の結果から, (1) 表層表現だけでもある程度文章構造解析が可能であること, (2) 文章構造解析結果をユーザに提示することで, ある程度の改訂支援が可能であることが分かった。

しかし, 現在の所, 以下のような問題点が存在する。

## 謝辞

本システムの構築に御協力いただいた中島靖氏(現三菱電機), 渡川智浩氏をはじめ, 改訂実験に御協力いただいた九工大情報工学部知能工学講座の皆様にご感謝いたします。

## 参考文献

- [1] 木下是雄: 理科系の作文技術, 中公新書 (1981).
- [2] 牛島和夫, 菅沼明, 富士通ソーシャルサイエンスラボラトリ: 推敲マニュアル(入門編・解説編), 岩波書店 (1993).
- [3] 住友金属工業: 日本語スベルバイザーユーザーズガイド, 住友金属工業 (1993).
- [4] 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).
- [5] 福本淳一: “筆者の主張に基づく日本語文章の構造化”, 情報処理学会自然言語処理研究会, Vol.78, No.15 (1990).
- [6] 黒橋禎夫, 長尾眞: “表層表現中の情報に基づく文章構造の自動抽出”, 自然言語処理, Vol.1, No.1 (1994).
- [7] 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, 電子情報通信学会技術研究報告, NLC92-43 (1992).
- [8] 中島靖, 渡川智浩, 中村順一, 吉田将: 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作成, 「自然言語処理における実動」シンポジウム論文集 (1993).
- [9] 野村浩郷: 自然言語処理の基礎技術, 電子情報通信学会 (1988).

表 5: 1人当りの改訂数

改訂の種類	教官	手作業	システム使用
記号に関する改訂	2.0	0.5	6.6
	1.6	3.4	14.2
語句に関する改訂	85.5	12.5	25.1
	69.2	84.7	54.0
文に関する改訂	19.0	0.8	3.3
	15.4	5.1	7.0
文間接続関係に関する改訂	15.5	0.8	10.6
	12.6	5.1	22.8
スタイル変更	1.5	0.0	0.8
	1.2	0.0	1.8
段落間関係に関する改訂	0.0	0.3	0.1
	0.0	1.7	0.2
全体	123.5	14.8	46.4
改訂指示, コメント	8.5	5.8	0.0

1. 現在のシステムでは, 単純なボタンマッチングだけで文の特徴抽出を行っているため, ボタンマッチングミスが原因で間違った特徴を抽出することがあること, また, 文章構造解析に有用な情報でも抽出できないもの(格情報等)があるといった問題がある。
2. 表層表現だけを用いているので, 文章構造の解析結果の曖昧さを排除することが難しい。
3. 文章構造の良し悪しに関する知識が無く, 『良い文章』に対する定義が曖昧なために, 改訂後の文章の評価が難しい。
4. 初心者には, どのように改訂すべきかを指示しないと, 現実問題として, より良い文章にはならない。改訂場所の指示だけでなく, 改訂方法のヒントが必要である。

1の問題に対処するために, 本システムでも形態素解析や格構造解析を行なうことが考えられる。しかし, これらの処理を行なうには, かなりの処理時間を犠牲にしなければならない。これについては今後の検討が必要である。

また, 2~4は本質的な問題であるため, 今後更に事例の調査を行い, 手がかりとなる表層表現の追加, 抽出機能の拡張, 効果的な改訂または改訂指示方法の考察を行なうことが必要である。