

作文添削システムの開発\*

3G-1

柿沼 秀和 伊藤 毅志 古郡 廷治†  
電気通信大学大学院 電気通信学研究科 情報工学専攻†

1 はじめに

本稿の目的は、作文の添削システムを開発することである。作文の添削には多大の労力と時間を必要とする。形式的な間違いを指摘したり、訂正するだけでも、教師の負担は大きい。

文章を書く能力の未発達な生徒は、作文上でいくつかの基本的な「誤り」をおかす。その誤りには、明白な「間違い表現」と、「不適切表現」とがある。教師は添削の過程で、前者の誤りは訂正し、後者の誤りは指摘した上で、書き換えを促すか、訂正案を示す。本稿で述べる作文添削システム、TENSAKUは、この教師の役目を果たすものである。

2 TENSAKU: 作文添削システム

文書作成支援システムは、何らかの目的で、文書をつくる上での手助けをする。従来、文書推敲支援システムでは、綴りと誤りの訂正や、文法、文体、内容のチェックなどの推敲システムを開発してきた。

たとえば、牛島らの「推敲」は技術的な論文の推敲を支援するシステムである。<sup>[1]</sup>高橋らのMAPLEは、大型汎用機のマニュアルの原稿を対象として、文書構造解析、辞書・文法を用いた日本語解析などの処理を行なった結果から、修正指針、総合評価点、評価などの情報を提示することで、推敲支援を行なっている。<sup>[2]</sup>また、鈴木らの開発したCRITACは、ワードプロセッサによって書かれた日本語文書に現れやすい誤りを調査し、辞書、校正知識などを用いて誤りの候補を指摘する。<sup>[3]</sup>

推敲システムは、すでに文章を書く能力を備えたユーザを想定したものである。したがって、システムは、文書中で誤りの候補は指摘するが、実際にその誤りを直すかどうかの最終的な判断はユーザにゆだねている。ここでは、英語のスペルチェックと同様、その主要な目的は“うっかりミス”の発見とその指摘にある。

一方、小学生は、作文能力そのものに問題をもっている。そのため、作文教育では、誤りの候補を指摘するだけでなく、文章を書く上での基本的な指導を必要とする。したがって、TENSAKUでは、小学生の書いた作文の解析を行い、作文上の間違い表現は修正し、不適切表現は指摘するとともに、分かりやすい言葉でその不適切性を説明する。

3 作文の分析

作文を添削するには、実際に小学生が作文でどのような間違いをおかしているかを知る必要がある。そこで、小学校高学年生の書いた作文<sup>1</sup>を対象に調査を行ってみたところ、約20項目の誤りのあることが判明した。

小学生の作文の誤りには、1文中の1単語にある単語レベルの誤り、2つ以上の単語のつながり、または、1文として見たときに、その文に異常のある文レベルの誤り、文章(作文)として見たときに、誤りであると判断できる文章レベルの誤りがある。

次の表は、それぞれのレベルの誤りを示したものである。ここで、添削欄の“修正”はTENSAKUで修正するもの、“指摘”は指摘対象とするものを示す(表中、第2欄の英文字のうち、大文字で示した項目は、現時点のTENSAKUで処理可能な誤りを示す)。

誤りのレベル		誤りの種類	添削
単語	A	当て字の使用	指摘
	B	送りがなの誤り	修正
	C	漢数字と算用数字の同時使用	修正
	d	活用誤り	修正
	e	同音漢字の誤用	修正
	f	同音ひらがなの誤用	修正
文	G	長すぎる文	指摘
	H	句点の誤用	修正
	I	読点の乱用	修正
	j	否定文の誤用	指摘
	k	助詞(が、は、を)の誤用	修正
	l	文のよじれ	指摘
	m	1文に1アイデア	指摘
	n	同意語の反復	指摘
文章	O	段落を変えていない会話文	修正
	P	接続詞の乱用	指摘
	Q	文末における誤り(文体)	指摘
	R	漢字の使用率の多少	指摘
	S	段落の構成(形式)	修正
	t	段落の構成(意味)	指摘
	u	指示詞の使い方	指摘

表: 作文上の誤り

以下に作文上の誤りの具体例をあげておく。ただし、文章レベルの誤りは、紙幅の関係で省略する。

- A 自分でも、なかなか、うまく出来たなと思った。
- B いく時も、備える時も、はいしました。
- C 五〇メートルだからたいへんでした。
- d あと卒業の練習も六年生といっしょに練習したりして楽しかったです。
- e のぼっている止中に空気を思いきりすつたらいい空気だと思っていっぱいすつた。
- f 運動会の日みんなのおどり お見た。
- G この間の台風24号の時、はくとお兄ちゃんは、安心して平気だと言っているのに、お母さんは、夜ねる時、ふとんの前に、ラジオとか、家のお金をひとつの袋の中に入れて、まくらもとにおいておいたのでびっくりしました。
- H あそんでた、はくは、いった、つかまらなかった。
- I 駅に、着いて、秋川まで、あるいて、十五分かかった。
- J とときき遊ばないこともあります。
- K わたしは、はじめて六年生をあてた。
- l わたしは、仲よしは、飯田さんと谷わきさんです。
- m それをのんで死んだらうと、はくは思ったけど今日は、ぜんぜんおもしろくない
- n 六級を受けたけど、六級はまだむりだった。

4 TENSAKU の概要

次の図はTENSAKUの概要を示すものである。TENSAKUは、前処理部(pre-processing)で作文を入力し、計算機可読文書(以後、作文ファイルと呼ぶ)を作成する。誤り検出部(error detection)に作文ファイルを送り、誤

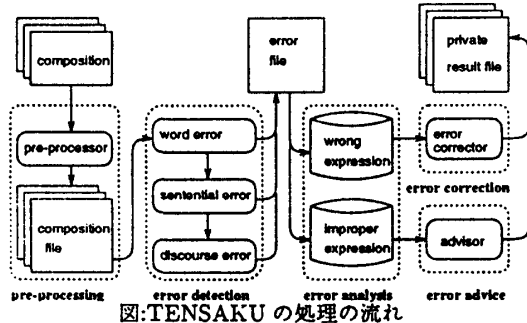
\*A Tool for Detecting and Correcting Errors in Compositions Written by Primary School Students

†Hidekazu Kakinuma, Takeshi Ito, Teiji Furugori

‡Department of Computer Science, University of Electro-Communications

<sup>1</sup>筆者らが小学校のときの作文集より、全122作文を調査した。

りを抽出する。抽出した誤りは誤り分析部 (error analysis) に出力し、誤りの傾向を調べる。最後に、間違い表現は誤り修正部 (error correction) で修正し、不適切表現は誤り指摘部 (error indication) で指摘する。



## 5 TENSAKU での誤りの処理

以下で、実例を用いて各部で行われる TENSAKU の処理を説明する。

### 5.1 前処理部 (pre-processing)

ここでは、作文ファイル作成のため、漢字コードの変換を行うとともに、TENSAKU の後の処理のために、句点ごとに区切って改行し、行番号を入れるなどの処理を行う。以下は、小学生の作文 (4 年生) を前処理した結果である。

題名: ドッジボール  
 1: わたしはドッジボールがこわかった。  
 2: でもわたしはがんばった。  
 3: わたしははげでばっかした。  
 4: ボールもとれなかった。  
 5: わたしは、はじめて六年生をあてた。  
 6: 私はここらの中でよかったともった。  
 7: でもわたしはまだときどきしながらはりまわっていました。  
 8: わたしは、「タイム。」といいました。  
 9: そしてわたしはボールをとろうとしたらだれかがとった。  
 10: わたしはボールをとろうとしたけど、あてられました。  
 ...以下略...

### 5.2 誤り検出部 (error detection)

単語、文レベルの誤りは、簡単な辞書を用い、1 文の部分文字列を抽出したものと辞書の内容とを比較することで検出する。

検出した誤りは次のようなリストにする。

(誤りのレベル (誤りの種類 行番号 誤りを含む文 誤りの開始場所 誤りの終了場所 添削法))

「誤りのレベル」という要素には、「単語」、「文」のいずれかが入り、「添削法」の要素には間違い表現ならば「修」、不適切表現ならば「指」が入る。

生成されたリストは誤り分析部に送る。誤りがない場合は nil とする。

実際に生成するリストは次のようなものである。これは誤りが 1 つの場合のリストである。誤りが複数ある場合には、このリストの要素が複数になる。

(「単語」  
 ("同音ひらがな" 6 "私はここらの中でよかったともった。"  
 26 28 "修"))

文章レベルで検出した誤りでは、種類によってリストの内容が変わる。ここでは、漢字の使用率に関する誤りの場合についてのリストの内容を示す。

(「漢字使用率」 個数 漢字の種類 使用率 (= 100 × 個数 / 文字数) 個数 / 種類)

前掲の作文からこのリストを生成すると、以下のようになる。

(「漢字使用率」 18 8 4.2 2.25)

### 5.3 誤り分析部 (error analysis)

誤りのリストは次に示すようなファイルに出力する。

(「単語」 ("同音ひらがな" 6 "私はここらの中でよかったともった。"  
 26 28 "修"))  
 ("文" nil)  
 ("文章" ("漢字使用率" 18 8 4.2 2.25 "指")  
 ("文頭" "わたし" 19 "指")  
 ("文体" "常体" "敬体" 7)  
 ("文体" "常体" "敬体" 8)  
 ("文体" "常体" "敬体" 10)  
 -略- "指"))

この結果から、全体でもっとも多い誤り (ここでは文体) を作文上の誤りの傾向とし、次のような文によって総合的な注意を与える。

【あなたの作文について】  
 作文を書くときには「～た・～った」と「～です・～ます」のどちらか 1 種類だけを使うように気をつけましょう

### 5.4 誤り修正部 (error correction)

ここでは、生成した誤りのファイル进行处理する。先の同音ひらがなのリストを例にすると、添削法が修正 ("修") なので、13 文字目にある "を" を "お" に直し、作文ファイルの 6 行目を次の正しい文と置き換える。

6: 私はここらの中でよかったとおもった。

次に、誤りの場所、直し方などを指示する。以下は、その結果である。

【まちがいの直しかた】  
 a) 場所: 6 行め  
 私はここらの中でよかったとおもった。  
 ↑  
 【まちがいの直しかた】  
 a) この "を" は "お" に直しましょう。

### 5.5 誤り指摘部 (error advice)

ここでは、表の添削欄が指摘になっている誤りを指摘する。不適切表現という言葉は分かりにくいので、ここでは「良くない所」と表示する。不適切表現には次のような指摘をする。

【良くない所】  
 c) 25 行中 19 行が「わたしは」という始まり方をしています。  
 【良くない所の直しかた】  
 c) 主語 (わたし) は書かなくても分かる場合があります。自分で読んでみて、いらなそう思ったところにある「わたしは」は省きましょう。

## 6 考察

TENSAKU は表層上の誤りを抽出する。しかし、当然、文の意味を考慮しないと判断不能の誤りもある。その点で、文法解析、意味解析も必要である。

TENSAKU は誤りを指摘するという「批判的」な教育を行っている。しかし、子供に共感を与えるような注意をしたり、良い表現をほめる機能を付加することも教育という観点からは必要である。このような機能をどうシステムにもたせていくかも、今後の重要な課題である。

## 参考文献

- [1] 牛島 和夫: "推敲", bit, vol23, No.1
- [2] 高橋 善文ら: "計算機マニュアル推敲・査読支援システム MAPLE の開発と運用", 情報処理学会論文誌, vol31, No.7 (1990)
- [3] 鈴木 恵美子ら: "日本語文書校正支援システムの設計と評価", 情報処理学会論文誌, vol30, No.11 (1989)