

# 科学的文章の推敲・校正を支援する 教育システムの構築

松 本 章 代

## 1. はじめに

著者が所属する情報科学科では、1年次の必修科目「初年次教育」において日本語の「科学的文章」の書き方を学生に指導している。ここでいう「科学的文章」とは、科学的な事柄についての文章を意味するのではなく、「伝えるべき事柄を正しくわかりやすく読み手に伝える文章」のことである。このような文章を書く力は、理系の学生のみならず、すべての大学生にとって身に付けるべき能力である。

文章作成指導のもっとも有効な手段は、担当教員によるきめ細かい添削指導であると考えられる。しかしながら、大人数を対象とした授業において添削指導を行うとなると、教員の労力は膨大なものとなる。しかも、添削指導の際には、一度読み直せば気が付きそうな不注意によるミスや、ルールを知ってさえいれば防げる誤りが多く目につく。添削する立場としては、せめてこのようなミスは無い状態で提出してほしいと願うが、自分自身が書いた文章を客観的に見直すことは難しく不備・不具合に気が付きにくいものである。ルールを守った文章が書けるようになるのにも、完全に身に付くまである程度トレーニングが必要である。また、学生の多くは他人に読んでもらうための文章を書くという意識が希薄である、ということも問題の一因である。

そこで「自分自身が書いた文章を客観的に見直す手助けをする」「科学的文章のルールを身に付けさせる」ための機能を有し、「推敲・校正の必要性を理解させ、他者に提出する文章は十分に吟味する癖をつけること」を目的とした、教育システム「TWSS (Technical Writing Support System)」を数年前から構築している。さらに平成22年度からは、これを実際の授業で運用している。本稿では、実装したシステムの機能とその授業実践について報告する。まず2章で関連研究について述べ、3章で実装したシステムの概要を説明する。4章で「初年次教育」における取り組みについて紹介し、5章で成果をまとめる。

## 2. 関連研究

文章を自動添削し、評価をユーザにフィードバックする教育システムとしては、米国 ETS が開発した「E-rater [1]」、石岡ら [2] による日本語小論文自動評価採点システム「Jess [3]」、日本語作文小論文検定協会による日本語の文章解析ソフト「森リン [4]」など、これまでも多数のソフトが開発されている。これら従来の文章添削システムは、文法の正しさ、語彙の豊富さ、文の長さ、漢字の量といった、表面的な指摘を行うものが主流である。これらは主に中高生の作文・小論文を対象としており、高等教育機関における科学技術論文の執筆指導を目的としたものではない。

一方、研究レベルでは技術文書を対象とした推敲支援ツールがいくつか開発されている。菅沼ら [5] は、マニュアルの執筆を想定し、読み手に誤解される文の検出を行っている。本システムが機械学習を用いて「わかりにくい文」を統計的に判断する [6] のに対し、菅沼らはヒューリスティックな理論に基づき判断を行う仕組みを提案している。また、稲積ら [7] は大学生の日本語文章力の育成を目的として、校正推敲支援や文章構造理解支援など 5 種類の支援ツールを開発している。これらのツールに備わっている、技術文書を書く上で順守すべきルールを指摘できる機能や、長文について係り受けの確認と修正を支援する機能は、我々が構築しているシステムの一部と類似している。ただし、本システムは品詞や主語・述語を色・形によって区別し、文章をより視覚的に意識させることができる。また、論理性の支援においては、稲積らのシステムがパラグラフライティングの観点から行われるのに対し、本システムは接続詞や接続助詞の見直しに重点を置いている。

## 3. システム構成

### 3.1 システム概要

TWSS は、「校正」を支援する「科学的文章のルールチェック機能」と、簡潔性・一義性の観点から「推敲」を支援する「わかりにくい文の指摘・可視化機能」、論理性の観点から「推敲」を支援する「全体の流れの可視化機能」の 3 つの機能から構成される。

なお、本システムは、ウェブアプリケーションである。インターネットにつながっていてウェブが閲覧できる環境なら、OS やブラウザに依存することなく利用可能である。ソフトウェアのインストールは（プラグインなども含め）一切不要である。

### 3.2 入力から出力までの流れ

TWSS を利用する際の流れは次のとおりである。システム画面を図 1 に示す。左側のテキストボックスは入力用、右側のテキストボックスは結果出力用となっている。

#### ① 科学的文章のルールチェック機能

- (1) 左のテキストボックスに文章を入力する。
- (2) 「基本ルールチェック」ボタンを押す。
- (3) 科学的文章を書くうえで順守すべきルールの違反チェックおよび意図が伝わりにくい文の検出が行われ、右のテキストボックスに結果が出力される。
  - ルール違反は該当する語が赤・橙字で表示される。(4) へ。
  - 意図が伝わりにくいと判定された文には「この文を可視化」ボタンが出現する。  
② へ。

- (4) 赤・橙字にカーソルを合わせると指摘内容がポップアップ表示される。赤字は要修正・橙字は要確認という意味である。

- (5) 左のテキストボックス内の入力文章を修正したら再び (2) へ。

#### ② わかりにくい文の指摘・可視化機能

- (1) 「この文を可視化」ボタンを押す。
- (2) 新しいタブが開き、その文を可視化した図および「意図が伝わりにくい」と判定された理由<sup>1</sup>が表示される。
- (3) 左のテキストボックス内の入力文章を修正したら再び (1) へ。

#### ③ 全体の流れの可視化機能

- (1) 「文書全体を可視化」ボタンを押す。
- (2) 新しいタブが開き、入力文書全体の流れを可視化した図が表示される。
- (3) 左のテキストボックス内の入力文章を修正したら再び (1) へ。

なお、①②③の順番で作業を進めなければならないわけではない。たとえばテキストボックスに文章を入力した後、「基本ルールチェック」ボタンを押さずに③に進むこともできるし、①で指摘された問題点が未解決の状態で②に進むことも可能である。

---

<sup>1</sup> 「主語と述語の距離が離れています。」「修飾句・節が多すぎます。」など。

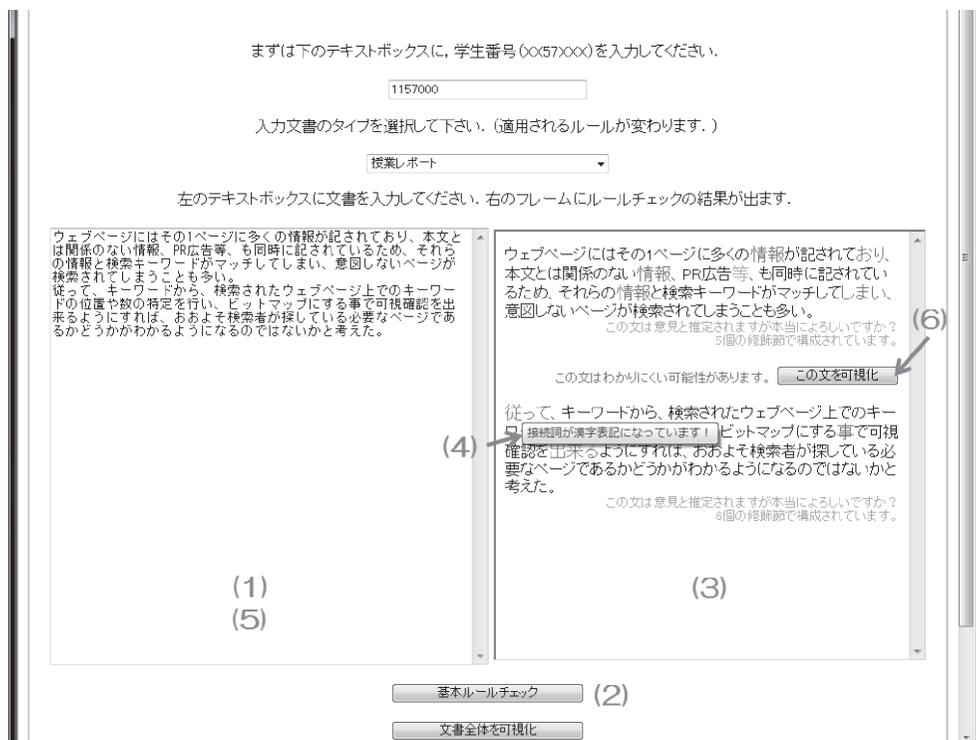


図 1. システム画面

### 3.3 3つの機能

#### ① 科学的文章のルールチェック機能

科学的文章には書き方のルールがある [8]。ただし、文書のタイプによって、そのルールは若干異なる。そこで TWSS では、文書のタイプに応じて、ルールに沿って書かれているかどうかをチェックする。そのため利用者は、あらかじめ「授業レポート」「科学技術論文」「その他（就職活動用エントリーシートなど）」から文書のタイプを選択しておく。チェックされるルールの種類は「科学技術論文」がもっとも多く（つまりチェックが厳しい）、「その他」はもっとも少ない。本システムでは文献 [8] [9] [10] を参考とし、様々なルールを採用している。チェックされる主なルール違反について「付録. チェック項目一覧」に示す。

続いて、わかりにくい文の判定について述べる。判定基準については、文を構成する単語数や主語の位置など、各文を形態素解析・構文解析した結果の情報に基づき、機械学習によって判別式を作成している。

## ② わかりにくい文の指摘・可視化機能

「わかりにくい」と判定された文は、図2のように文単位で可視化される。矢印はエッジといい、修飾—被修飾関係を表す。格助詞（「が」「を」「に」「で」など）・係（副）助詞（「は」など）によって動詞を修飾する関係は、特に重要とみなし、エッジを目立つ赤色にしている。また、接続助詞は論理的关系を決定づけるため、注目しやすいよう黄色にしている。なお、四角で囲まれた語は、主語と述語である。

この図から利用者が読み取るべきことと行うべき作業について、次に述べる。

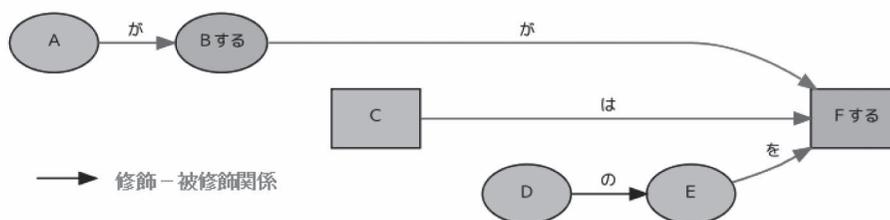


図2. 文の可視化

### (1) 語間の修飾—被修飾関係とその距離

利用者は、可視化された図を見て、語間の関係が正しいか（本システムの解釈と利用者の意図とが同じになっているか）否かを確認する。システムに誤った解釈をされるようであれば、複数の意味に解釈できる文になってしまっている可能性が高い。そのような場合は、利用者の意図と一致するよう書き変えるべきである。

また、長いエッジのある文章、すなわち長い修飾節を間に挟んだ文章は「逆茂木型 [11]」と呼ばれ、読みにくい文の典型として知られている。文節を入れ替えて短いエッジにできないかどうか、あるいは文を切るといった対応を検討すべきである。

図3は「金色の雨が五月の明るい太陽の下で輝く若葉に降りそそぐ。」（出典：文献 [9]）を本システムにより可視化した図である。しかしながらこの図の係り受け関係は誤っている。これは利用者の意図どおりに構文解析が行われなかったことを意味する。もしもこの文が意図どおりに解釈された場合（図4）は、「雨」と「降りそそぐ」が非常に長いエッジで結ばれることになる。そこで文を「五月の明るい太陽の下で輝く若葉に金色の雨が降りそそぐ。」と修正すると、エッジの長さが総合的に短くなる（図5）。

### (2) 一文を構成する句・節の数

1文を構成する句・節の数が多いほど文の構成は複雑になり、わかりにくさの要因となる。しかし、普段文章を書くときには句や節を意識することはなく、判断が難しい。そこで、こ

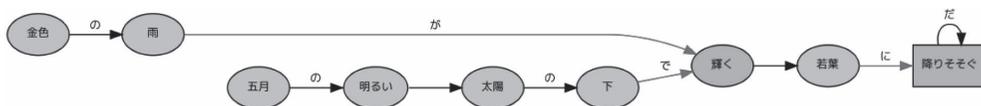


図 3. 「金色の雨が五月の明るい太陽の下で輝く若葉に降りそそいだ」を可視化

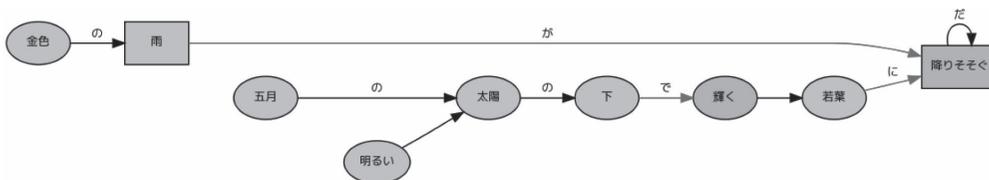


図 4. 図 3 が正しく解釈された場合

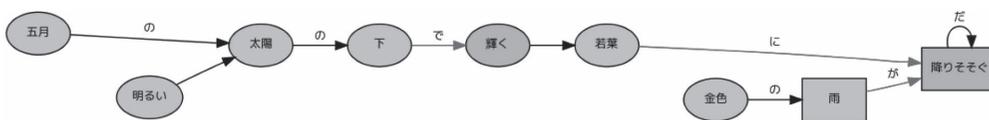


図 5. 「五月の明るい太陽の下で輝く若葉に金色の雨が降りそそいだ」を可視化

の図では動詞を水色で表現している。大雑把に言って、句や節の多い文章は文中の動詞が多い。利用者は、水色の語が多いと感じた場合には文を分割する。

### (3) 主語と述語の関係

文章を書くうえで、主語と述語の関係を意識することは非常に重要である。そこで、利用者は以下の 4 点についてチェックを行い、必要に応じて文章を修正する。

- 主語があるか
 

日本語は主語を省略することができる。しかし、主語の省略は文章をわかりにくくする要因になる。特に、直前の文と主語が変わる場合には、主語は省略しない方がよい。
- 主語に対し述語の表現は適切か
 

主語と述語がねじれていないか確認する。
- 主語や述語は意図した語が選ばれ関係づけられているか
 

意図どおりでなければ、(1)と同様に、図が正しくなるように文章の構成を修正する。
- 主語と述語の距離が離れすぎていないか
 

離れすぎている場合は、(1)と同様に、エッジが短くなるように修正する。

③ 全体の流れの可視化機能

文書の論理性が損なわれる要因には

- 述べる順序が不自然
- 接続詞の省略
- 文の欠落（論理の飛躍）
- 接続詞の不適切な使用

などがあげられる。

本システムでは、「余計な修飾表現は無い方が、話の流れに集中し易く、文と文の関係を見直す作業の支援につながる」という我々の仮説に基づき、文書全体のあらすじを可視化することにより、論理展開のチェックを支援する。

文書全体を可視化した例を図6に示す。利用者は、システムに表示された図を見て、筋がとおっているかどうかの確認を行う。特に、黄色で示された接続詞・接続助詞に着目する。論理の飛躍はないか、文の順序は適切か、接続詞の不足はないか、に注意しながらチェックしていく。

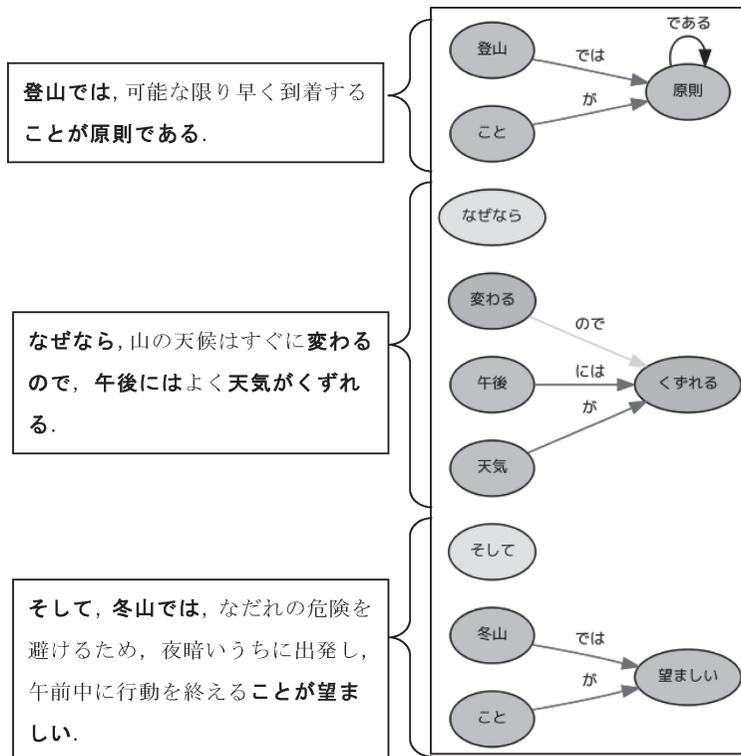


図6. 全体の可視化の例

#### 4. 授業実践

本システムは、情報科学科1年次必修科目「初年次教育」において実際に導入されている。2013年度における本システムを利用した授業について述べる。

##### 4.1 2013年度の取り組み

全15回の授業のうち、本システムの実習を伴う授業は2回分（第6回・第12回）である。簡潔性・一義性・論理性については、その前（第3回～5回）の授業においてあらかじめそれぞれ講義を実施している。

第6回は「推敲と校正」と題し、推敲と校正の違いについて説明した後、主に「推敲」についての実習を行う。学生は、簡潔性・一義性・論理性の観点から、本システムを利用して文章の修正問題を解く。

第12回は「科学技術系レポートの構成と基本ルール」という題で、主に「校正」の実習を行う。レポートの作成において守るべき基本ルール<sup>2</sup>を示した後、本システムを利用して文章の修正作業を行ってもらう。

また、この授業の最終レポートについても、提出前にTWSSを使用するよう指導している。

##### 4.2 学生に与えるべき注意

なお、本システムを学生に利用させるにあたり、与えるべき注意が2点ある。以下に示す。

- 本システムで表示された警告は、必ずよく検討すること。ただし、指示が明らかに不適切だと感じる場合は従わなくてかまわない。コンピュータによる判断には限界がある。
- 本システムで不具合が検出されなかったからといって「完璧な文」であるとは限らない。本システムでは、語の「意味」を考慮した判定を行うことはできない。また、構文のおかしさ（型が正しくない文）を検出できない。主語－目的語－述語の関係がチグハグである文ならば、可視化された図はヒントになる。しかし「文の型がおかしい」という警告メッセージが出力されるわけではないので、気がつくことが難しいケースもある。

---

<sup>2</sup> 本システムでは対応していないルール（「A4用紙・横書き」が原則、ルーズリーフ不可、ホチキスで左上をとめる、など）も含む

## 5. むすび

本論文では、開発した科学的文章を推敲・校正する教育システムの機能について紹介し、授業における取り組みを報告した。

現在、「全体の流れの可視化機能」に、論理的に問題のある個所を自動検出し指摘する機能を付加することを目指し、非論理的な接続表現の具体的な検出手法を検討している。また今後は、本システムが論理性的の推敲の手段として有効であることや、システムを実際の授業に導入した場合の効果について、検証していく予定である。

なお、本システムの URL は次のとおりである。どなたでも自由に利用いただいて構わない。

<http://mmt1.cs.tohoku-gakuin.ac.jp/twss/>

## 謝辞

本研究は JSPS 科研費（若手 B，課題番号 24700906）の助成を受けている。

## 参考文献

- [1] Attali, Y. and Burstein, J.: Automated essay scoring with e-rater v. 2, *Journal of Technology, Learning, and Assessment*, Vol. 4, No. 3 (2006).
- [2] 石岡恒憲, 亀田雅之: コンピュータによる小論文の自動採点システム Jess の試作, *計算機統計学*, Vol. 16, No. 1, pp.3-18 (2003).
- [3] Jess, <http://coca.rd.dnc.ac.jp/jess/>.
- [4] 森リン, <http://www.mori7.info/moririn/index.php>.
- [5] 菅沼 明, 小野貴博: 文章推敲支援における読み手に誤解される文の抽出, *情報処理学会研究報告*, 2007-DD-61, Vol. 2007, pp. 31-38 (2007).
- [6] 鈴木雅人, 松本章代, 田中大輔, 山田未央佳, 山田 翔, 北越大輔: 理工系学生を対象とした文章作成能力向上のための支援システム, *東京工業高等専門学校研究報告書*, No. 40 (1), pp. 59-62 (2009.1).
- [7] 稲積宏誠, 大野博之, 竹内純人, 大久保麻里子, 又平恵美子: ICT を活用した日本語文章力育成への取り組み, *情報処理学会研究報告*, Vol. 2011-CE-109, No. 9 (2011.3).
- [8] 中島利勝, 塚本真也: 知的な科学・技術文書の書き方, コロナ社 (1996).
- [9] 阿部圭一: 明文術—伝わる日本語の書きかた, NTT 出版 (2006).
- [10] 戸田山和久: 論文の教室—レポートから卒論まで, 日本放送出版協会 (2002).
- [11] 木下是雄: 理科系の作文技術, 中公新書 (1981).

付録. チェック項目一覧 (科学技術論文の場合)

- 1 文中に同一単語が頻出している
- 助詞「の」が連続している (文がわかりにくくなる)
- 助詞「が」「を」「に」について 1 文中に同じものが複数存在する (文がわかりにくくなる)
- 1 文中に同一助詞が頻出している
- 同じ文末表現が続いている
- 体言止めが使われている
- 連用形が多用されている (文がわかりにくくなる)<sup>3</sup>
- 接続助詞「が」が使われている (文がわかりにくくなる)<sup>4</sup>
- ら抜き表現が使われている
- 当て字が使われている (例: 容易い, 相応しい, など)
- 接続詞が漢字表記になっている (例: 従って, 然し, 故に, など)
- 形式名詞<sup>5</sup>が漢字表記になっている (例: 検索した所, 検討した上で, など)
- 補助動詞<sup>6</sup>が漢字表記になっている (例: 計算して見る, 減少して行く, など)
- 副詞には漢字で書くべきものとひらがなで書くべきものがある (例: 予め, 全て, 先ず, など)
- 1 文中に修飾語 (形容詞・副詞) が使われすぎている
- 1 文中に修飾節が 2 つ以上ある
- 口語的表現が使われている (例: きちんと, そんな, こんな, など)
- 「など」「たり (だり)」が使われている (複数例示されていないと文の意味があいまいになる)
- 意見かどうか (意見と推定される文に対して, 本当に「意見」でよいのか確認を促す)
- 常体で統一されていない ☆
- 読点が「,」ではなく「、」になっている ☆
- 一人称単数の主語が使われている ☆
- 常用漢字ではない ☆

☆印は入力文書のタイプが「科学技術論文」のときのみ適用される

<sup>34</sup> 連用形の多用や接続助詞「が」の使用を避けた方がよい理由については, 参考文献 [10] に詳しい.

<sup>5</sup> 本来の漢字の意味を持たない例のような名詞のこと

<sup>6</sup> 本来の漢字の意味を持たない例のような動詞のこと