

読点の打ち方と語の並べ方の競合回避による 日本語文章の推敲支援手法

須藤 崇志[†] 丸山 広[†] 鈴木 健之[†] 中村 太一[†]

東京工科大学[†]

1. はじめに

論理的で分かりやすい技術文章を書く能力は、社会人にとって必要である。乾らは、ビジネス文章を対象に推敲作業を分析した結果、推敲理由は 6 つに大別でき、係り受け関係の分かりにくい文の修正方法には、「読点の追加」「修飾語の並べ替え」「置き換え」「削除」の 4 つがあることを発見した[1]。

本研究では、「技術文章初学者の文章の分かりやすさを向上させる推敲支援手法の確立」を目的とする。

技術文章が分かりにくい原因は、ビジネス文章と同じであると考え、係り受け関係の分かりにくい文を修正する方法の 51%は、読点の追加と修飾語の並べ替えであるという乾らの分析結果から、読点と語順の作文規則[2]を基に推敲支援に取り組む。係り受け関係の分かりにくい文を検出するルールとルールの競合による過検出を回避するルールを設定し、従来研究[3]で提案したルールと併せ、当該文を指摘する推敲支援手法を提案する。

2. 係り受け関係の分かりにくい文と修正方法

係り受け関係の複雑な文は分かりにくい。例えば「太郎は慌てて逃げる次郎を追った」という文は「太郎は」「慌てて逃げる次郎を」2つの節が「追った」を修飾している。しかし、「太郎は」の節は、「慌てて」の節にも係っている。このため、「慌てて」の節の係り受け関係が複雑なため文意が一意に定まらない。

係り受け関係の複雑さは、節を長い順に並べ替えたり（以下、語順）、短い節が長い節の前にある場合（以下、逆順）や長い節が隣接している場合（以下、境界）には間に読点を追加したりすることで解消できるとされる。しかし、並び替えと読点の追加は同時に適用できない場合がある。並び替えと逆順の読点を同時に適用

すると、「慌てて逃げる次郎を太郎は、追いかけた」となり、不要な読点を追加してしまうため、ルールの取捨選択が必要である。

3. 分かりにくい文の指摘の流れ

分かりにくい文の指摘の流れを図 1 に示す。

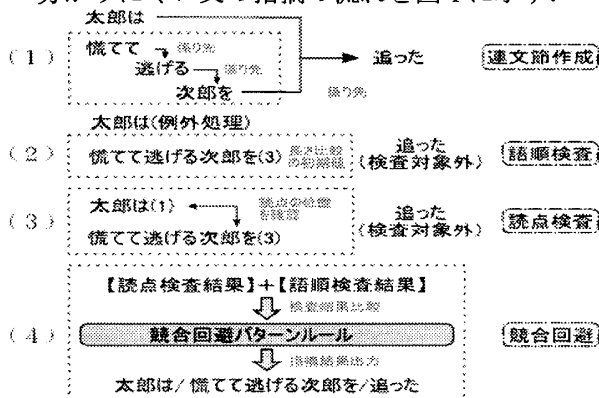


図 1 分かりにくい文の指摘の流れ

(1)連文節作成では、文節を結合し、連文節にする。連文節作成の条件は、隣接した 2 文節間の係り受け関係が 1 対 1 である場合とした。

以降の語順検査や読点検査では、連文節の長さや位置を見る。長さは、2 文節長以下を「短い」、3 文節長以上を「長い」とした。

(2)語順検査では、主題表現・接続詞・時刻表現を含む連文節は例外で対象外とし、係り先が同じ連文節群で短い連文節の後ろに置かれている長い連文節を検出する。

(3)読点検査では、検出ルールを以下の A・B とする。

- A. 短長の順で連文節が並んでいる場合、間に読点を打っていない連文節を検出する。
- B. 長い連文節が隣接している場合、間に読点を打っていない連文節を検出する。

(4)競合回避では、語順検査結果と読点検査結果を文単位・連文節単位で比較し、検出の状態や分岐要因の組み合わせに対応した、文単位 4 種・連文節単位 16 種のパターンルールを適用し、最終的な指摘結果を出力する。表 1 にパターンルールの一部を記載する。

The Methodology of Polish Support for Japanese Text by Competition Avoidance of How to Strike Comma and How to Set Word.

[†]Takashi SUDOU [†]Hiroshi MARUYAMA [†]Kenshi SUZUKI [†]Taichi NAKAMURA; Tokyo University of Technology.

表1 競合回避パターンルール (抜粋)

比較単位	検出の状態	分岐要因	結果
文	読点検査のみ指摘		検査結果で指摘
文	語順検査のみ指摘		連文節単位で比較処理
文	両検査で指摘		連文節単位で比較処理
連文節	語順と逆順	例外処理あり	語順で指摘
連文節	語順と逆順	例外処理なし	逆順で指摘
連文節	語順と境界		指摘内容を統合

4. 検証方法

4.1. 検証対象の文章と検証手順

本提案手法が文章の分かりやすさ向上に寄与することを、東京工科大学コンピュータサイエンス学部システム系実験Ⅱのレポートを対象に検証した。提出されるレポートは、実験支援システム[4]で形式がチェックされている。提案手法検証用システム（以下、システム）で、形式がチェック済みのレポートに対して分かりにくい文を指摘する。

検証手順は、最初に実験支援システムに保存された文章を句点基準で1文に分割し、日本語係り受け解析器「CaboCha」[5]で、係り受け関係と品詞を同定する。最後に2で挙げた手法で分かりにくい文を指摘し、修正を促す。

4.2. 得られた結果

24名の学生がシステムを計123回利用した。その結果、システムには5759文が送られ、分かりにくい原因を1997箇所指摘した。システムの利用ログから修正後に指摘が無くなった文を50組得た。

5. 評価

5.1. 評価実験方法

教員1名と実験講師1名による、利用ログから抽出した推敲前後の文50組の分かりやすさの比較結果をもとに評価した。

5.2. 評価実験結果

評価対象50文の評価結果を表2に示す。

表2 分かりやすさの比較評価結果

評価者	評価	文数	割合
教員	修正後が良い	28文	56.00%
	修正前が良い	7文	14.00%
	修正前後で変わらない	15文	30.00%
実験講師	修正後が良い	9文	18.00%
	修正前が良い	6文	12.00%
	修正前後で変わらない	35文	70.00%

6. 考察

表2から、教員が「修正後が良い」と評価した文

は、全体の56.00%であり、実験講師が「修正後が良い」と評価した文は、18.00%であった。いずれも、「評価前が良い」より高く、教員の評価では、「修正後が良い」と評価した文が50%前後に達した。以上の結果から、学生の文章には、効果があることを確認した。

また、教員と実験講師に「修正前が良い」「修正前後で変わらない」と評価した理由についてアンケートを実施した結果、その理由は、「語句が違うだけで変わらない」「不要な読点がある」の2つに大別できた。語句が違うだけで変わらない原因は、指摘されている語を削除したり、指摘されない語に書き換えたりといった指摘外の修正をすることで指摘をなくした場合で、文の分かりやすさに変化はない文だった。不要な読点の原因は、文が日本語として正しくないため係り受け同定を誤り、文の検査に誤りが生じて不要な読点が追加された場合と、システムの指摘は正しいが元より読点が多い文や短い文は読むときに冗長に感じる場合の2つだった。日本語として正しくない文は、32文あり全体の64.00%だった。

7. おわりに

本研究では、文の分かりやすさを向上させる手法として、長い修飾語を前に置くルールと逆順や境界には間に読点を追加するルールを用いた。本手法を実装したシステムを学生のレポート作成時に使用してもらうことで検証した結果、学生の作成した文章では、「修正後が良い」という評価が最高56%までに達し、両名の評価でも「修正前が良い」という評価を上回った。しかし、指摘外の修正により指摘が本来の効果を発揮しなかった場合や不要な読点が追加される場合があることが分かった。今後は、指摘に沿った修正ができるような指摘方法の工夫や、日本語として正しくない文を正しくするための校正支援や、文の状態に応じて読点を削除する指摘に取り組み、技術文章初学者の文章の分かりやすさを向上させる推敲支援手法の確立を目指す。

参考文献

- [1] 乾裕子, 岡田直之: 長い文は常にわかりにくいかわかりにくさの要因とその依存関係~, 情報処理学会研究報告, Vol.2000, No.11, pp.63-70(2000)
- [2] 本田勝一: 日本語の作文技術, 朝日新聞社, (1982)
- [3] 大長達也, 丸山広, 中村太一: 係り受け構造を用いた読みやすい文章への推敲支援", 情報処理学会第69回(平成19年)全国大会, 6Q-6, pp. 2-433 - 2-434(2007)
- [4] 丸山広, 中村正太, 高橋知七, 中村太一: テキストマイニング技術の活用に向けた実験支援システム", 情報処理学会第68回(平成18年)全国大会, 5V-7, pp.4-639-4-640(2006)
- [5] 工藤拓, 松本裕治: チャンギングの段階適用による日本語係り受け解析, 情報処理学会論文誌, Vol.43, No.6, pp. 1834-1842(2002)