

文分割における連用中止表現の扱い

5 Q-9

武石 英二 林 良彦

NTT情報通信処理研究所

1. はじめに

マニュアル文書を対象とした日本文推敲支援システム[1]の研究開発を進めている。これは文章中の推敲すべき表現(不適性表現)を検出するとともに、それに対する書換え候補を提示するシステムである。マニュアルに頻出する不適性表現の1つに長文があり、上記システムにおいては、形態素解析情報に基づく文分割機能の検討を進めている[2]。

従来の文分割方式では、基本的に①分割する文節(以下、分割点と呼ぶ)の接続形式と分割時に挿入する接続詞を一対一に対応させる、②分割点を終止形に変換する、という方式[3]が多い。このとき連用中止表現や接続助詞「て」による接続(以下、本稿では両者を併せて連用中止表現と呼ぶ)は、通常並列接続として扱われるため、上記のような方式では「並列の接続詞の挿入(または接続詞なし)、分割点文節の終止形への書換え」という分割方法を採用と考えられる。

しかし、連用中止表現はその前後に特定の意味関係(以下、連用中止表現の意味用法という)を担うことがあることが指摘されている[4]。また実際にマニュアル文中の連用中止表現を収集した結果でも、単なる並列接続とみなし得ないものが比較的多く存在する(図1①)。さらに連用中止表現は、述語に付加される付属表現(時制表現や様相表現)が省略された表現形式であるため[5]、分割点文節を終止形に変換するという方法では有効な分割文候補を生成することができない場合がある(図1②、③)。

本稿では、文分割の観点から連用中止表現の意味用法を

整理するとともに、表現の形態的特徴と意味用法、分割時の接続詞、連用中止文節の形態素操作との関係に着目した連用中止表現を分割点とする文の分割方式について述べる。

2. 連用中止表現の意味用法と分割パターン

連用中止表現がどのような意味用法を担い得るのかについては、すでにいくつかの分析がなされている[5,6]。これらを参考にしながら、マニュアル文(2種類、314文)中の連用中止表現86表現を試料として、

- ・連用中止表現の他の接続表現への変換
 - ・分割文の作成
- を机上で行い、連用中止表現の意味用法の特徴についての考察を行った。

2.1. 意味用法分類

接続表現への変換、文分割の調査に基づき、表1に示す意味用法分類を設定した。ただし、「~」は連用中止表現を表しており、また「-」は分割することが不適切なことを示している。

表1. 意味用法分類

用法	接続表現への変換	分割時の接続詞
対比	~, また	また
添加	~, そして	そして
原因	~ので、	従って/そのため
方法	~ことによって、	——
根拠	~ことによって、	それによって

2.2. 意味用法候補

(1) 表現形態の分類

まず、連用中止表現を表現形態(連用中止表現を構成する自立部と付属部の組み合わせ)によって以下のように分類する(以下、連用中止分類という)。

- ①し型: 動作性動詞の連用形
- ②して型: 動作性動詞の連用形+て
- ③してて型: 動作性動詞の連用形+ていて
- ④でき型: 連用形+でき(可能)
- ⑤あり型: 状態動詞の連用形
- ⑥く型: 形容詞の連用形(+て)
- ⑥で型: 「だ」の連用形
- ⑦ており型: 動作性動詞の連用形+ており
- ⑧であり型: 「である」の連用形

次に連用中止表現を受ける述語を上記の連用中止分類との同一性という観点から以下のように分類する(以下、述語分類という)。

- ①する型: 動作性動詞で②、③以外
- ②ている型: 動作性動詞+ている
- ③できる型: 動作性動詞+できる(可能)

① Aシステムでは、ソフトウェア開発保守工程それぞれで必要となる各種ソフトウェアを揃えており、生産性の高いソフトウェア開発ができます。

→ Aシステムでは、ソフトウェア開発処理工程それぞれで必要となる各種ソフトウェアを揃えています。
[従って]生産性の高いソフトウェア開発ができます。

② Aシステムでは時代の動向を的確に把握分析し、ハードウェアとソフトウェアの計画的な開発を進めています。

→ Aシステムでは時代の動向を的確に把握分析しています。
[それによって]ハードウェアとソフトウェアの計画的な開発を進めています。

③ ハードウェア装置の故障を早期に発見し、エラー箇所の訂正などに対応することができます。

→ ハードウェア装置の故障を早期に発見することができます。そして、エラー箇所の訂正などに対応することができます。

~~~~: 連用中止表現      ——: 分割時に補うべき表現

— 図1. マニュアルにおける連用中止表現例 —

- ④ある型：状態動詞
- ⑤形容型：形容詞
- ⑥ダ文型：ダ文

(2) 意味用法候補ルール

連用中止表現の意味用法は、連用中止分類と述語分類の組み合わせと密接な関連を持っており、以下のような特徴がある。

- ・ 可能の様相をもつ述語に接続する連用中止表現では、原因の意味用法を担うことが多い。：これは可能の様相がその前件に原因／理由となる記述を要求することが多いことによるものだと考えられる。
- ・ 「で、であり型」の連用中止表現では対比や添加の意味用法を担うことが多い。
- ・ 「し型→する型／ており型→ている型」等の同一形態の述語の接続では対比や添加の意味用法を担うことが多い。：これが連用中止表現の最も基本的な文型である[3, 5]ことによるものと考えられる。

そこで連用中止分類と述語分類の組み合わせから連用中止表現の意味用法候補を決定する(例を表2に示す)。これは、各組み合わせにおいて意味用法を一意に決定するのではなく、複数の意味用法に順位付きで限定するものである。ただし、?は該当する文が資料中に存在せず、かつ日本語として不適切だと考えられるもの[5]である。

表2. 意味用法候補ルール例

|      | する型   | ている型  | できる型  | ... |
|------|-------|-------|-------|-----|
| し型   | ①添 ②原 | ①添 ②根 | ①添 ②根 | ... |
| :    | :     | :     | :     | ... |
| あり型  | ①添 ②原 | ①添 ②根 | ①原 ②添 | ... |
| :    | :     | :     | :     | ... |
| であり型 | ?     | ①添    | ①添 ②原 | ... |
| ており型 | ?     | ①根 ②添 | ①原    | ... |

2.3. 分割パターン

既に指摘した通り連用中止表現を含む文の分割では、付属表現の削除／付加等を必要とすることが多い(図1②、③)。また同一の意味用法をもつ接続詞の使い分け(従って／そのため等)を決定することは困難であるが、「で型、であり型」が添加の意味用法を持つ場合には、接続詞を挿入すると冗長になるといった傾向が見られる。

そこで、連用中止分類、述語分類及び意味用法によって、接続詞、分割点の形態素操作(以下、分割パターンという)を決定し(表3)、分割文候補を生成する。また、2.2で得た用法が複数個存在する場合は、複数の分割パターンによって分割し、分割文候補として出力する。

表3. 接続詞・形態素操作ルール例

| 分割点  | 文末   | 用法 | 接続詞  | 形態素操作        |
|------|------|----|------|--------------|
| し型   | できる型 | 添加 | そして  | +できる         |
|      |      | 原因 | そのため | +ている         |
|      |      | :  | :    | :            |
| ており型 | -    | 原因 | 従って  | -ており<br>+ている |
|      |      | :  | :    | :            |
|      |      | :  | :    | :            |
| であり  | -    | 添加 | -    | -である<br>+です  |

3. 評価と考察

(1) 評価

連用中止意味用法判定部、分割パターン生成部をコーディングして評価を行った。その際、文中に明示された接続詞についても一部考慮した。例えば連用中止表現の直後に接続詞が明示された場合には接続詞を挿入しない、文頭に接続詞が明示された場合には同一接続詞の接続詞を避けるなどである。評価の対象は、①ルール作成に用いたもの86表現のうち形態素解析の失敗を除く79表現、②ルール作成に用いたものと異なる75表現を用いた。意味用法の正解を予め人手で付与したものを用いた。

意味用法限定ルールについては、

①の場合) 1位: 46, 2位: 24, 候補なし: 9

②の場合) 1位: 54, 2位: 7, 候補なし: 14

であり、約85%に対して2位以内に有効な候補を出力することができる。また分割文候補についても、正しい意味用法が付与された文に対しては、①の場合で53文中43文、②の場合で45文中35文に対して有効な候補を出力することができる。

(2) 考察

上記の評価をもとに本方式の改善点について考察する。

・ 副詞的な表現への対処

形態的には連用中止であるが、構文的には副詞的な役割を果たしているものがある。そのような表現は固定的な表現(例えば「～に対応して」)が多いため、原則として予め辞書に登録している。しかし辞書に未登録の表現は連用中止として認定され、不適切な用法、分割文候補を出力してしまう。このうち、「～をはじめ」のようにマニュアル文ではほぼ副詞的に使用されると考えられるものについては、個別に対処することが可能だと考えられる。

・ 表現形態分類の詳細化

本稿で述べた表現形態分類は、ほぼ文節を構成する単語の品詞によるものである。さらに付属表現の意味による分類や頻出する動詞がマニュアルにおいてどのように使用されているかを整理することによって表現形態分類を詳細化することが考えられる。

・ 分割可否条件の設定

連用中止を含む文に限った問題ではないが、分割後の2文のバランス等を考慮していないため、不自然な分割候補を出力する場合がある。文脈的な情報を用いて分割する可否かについての判定についても検討する必要がある。

4. おわりに

マニュアル文中の連用中止表現の分析に基づいて、表現の形態的特徴に基づく意味用法候補、分割パターンの決定方法について述べ、その有効性を示した。今後は3で述べた考察を基に本方式の拡張について検討を進める予定である。なお2.2で述べた意味用法候補ルールは、推敲支援システムにおける連用中止の推敲機能(接続表現への変換: 例「～し→～することによって」)にも適用している。

[参考文献]

- [1]林, 菊井: 日本文推敲支援システムにおける書換え支援機能、第4回人工知能学会全国大会論文集、15-6、1990
- [2]武石, 林: 複文における分割点の特定について、第4回人工知能学会全国大会論文集、9-6、1990
- [3]信国: 自然言語における長文分割方式、第39回情処全国大会、2F-1、1989
- [4]寺村: 日本語の文法、国立国語研究所編、1980
- [5]三上: 現代語法新説、くろしお出版、1972
- [6]長尾他: 科学技術論文における並列句とその解析、情処研究会、NL-36-4、1983