

言語部品表現を用いた推敲支援

上田 良寛 水梨 豪* 竹田 幸史

富士ゼロックス(株) システム・コミュニケーション研究所

(* 現在、ATR音声翻訳通信研究所)

一般に、推敲とは、テキストに対して何度も書換えを行い、その効果を見ながら最適な表現を選択するプロセスである。推敲支援システムは多くの研究がなされているが、このようなプロセスをサポートするものはない。この推敲のプロセスの支援を行うためには、テキストに対する変更が容易に行えること、その変更の効果を書き手が知ることができる必要がある。このための枠組みとして、言語部品表現を提案する。言語部品表現は、文書の論理構造をテキストまで拡張したもので、これにより、文の句や節、単語単位での編集を容易にする。

Text Refinement Support System Based on Language-Parts Expression

Yoshihiro Ueda, Suguru Mizunashi* and Koji Takeda

Systems & Communications Lab., Fuji Xerox Co., Ltd.

(* currently belongs to ATR Interpreting Communication Research Labs.)

Text refinement is a process of repetitive modification of text and examining the effect of the modification. In the course of the repetitive modification, the writer selects the best expression s/he wants. Though many refinement support systems are studied and developed, they hardly support such a process. To support the refinement process, text must be easily modified and the writer must be able to know the effect of the modification. For this purpose, here we propose an architecture called "language-parts expression," which is an extended model of the logical structure of documents to text. The user can modify the text by its words, phrases and clauses represented as language-parts.

1. はじめに

推敲支援システムは、既に種々の研究がなされており、推敲支援の機能をもったエディタ等も発売されている。しかし、これらのもつ機能を見ると、

誤字脱字

文体、表記の揺れ

推奨できない文構成(受動態等)

意味の分かりにくい文

長文、複雑な構成の文、曖昧な文

など、悪文を指摘し、その書換えを支援するものである。これらのシステムの狙いは、テキスト品質を、「受け入れ難い品質レベル」から、「最低限の品質レベル」への底上げを支援するものといえる。

我々は、推敲を、「文章を書く人が、自分が表現したいと考えていること(意図)を、最も良く表現する文章を作り出すプロセス」ととらえ、このプロセスを支援することをねらいとしている。この意味では、これまでの推敲支援のねらいとは異なり、「さらに高い品質」を作りこむことを支援するものということができる。

文章によって「表現されている内容」は、文章の読み手がその文章を読むことによって初めて意味をなす。実際はこのように「表現されている内容」は、固定的にあるわけではなく、読み手によって変わるものである。それでもなお、推敲のプロセスにおいては、一般的な読み手を想定した上で、読み手の立場に立って文章を見直すことが要求される。読み手の立場に立つこと自体困難な場合が多い。

ここで述べる推敲支援システム X-ACT (Xerox's Assistant for Customizing Text) は、このようなプロセスをサポートすることを目的として研究を進めているものである。

ユーザがテキストに変更を加えるたびに、その変更がテキストに与える意味や効果、すなわち、読み手に与える意味や効果を知らせる。ユーザはそれを見て、さらに変更を加えるかど

うかを決める。この変更と効果の通知の繰り返しの中から、ユーザは最も良い表現、すなわち、「自分が表現したいと思っていること、読み手に伝えたい内容を、最も良く表現する文章」を選択する。

変更が容易にでき、その効果を反映させることができるようにするため、我々は、テキストを文字列として保持するのではなく、構造をもったものとして保持する方法を採用した。これを言語部品表現と呼ぶ。

本報告では、このためのアーキテクチャである言語部品表現について簡単に説明し、推敲支援機能について考える。

2. 言語部品表現と推敲支援機能

文書は表題から始まり、章、節等の論理構造をもっている。通常のエディタを用いる場合、論理構造は文字列の中で暗黙に保持されるだけだが、最近では、アウトラインプロセッサやSGMLエディタなど、文書の論理構造を保持し、編集可能にしたものが現れている。

通常、文書の論理構造は章、節の単位で止まっており、テキストはその内容として扱われ、それ以上構造化はなされない。言語部品は、文書の論理構造を、テキスト内部まで拡張したものである(図1)。構造化を、パラグラフ、文、文の内部(句、節、単語)まで進める。これにより、単語、句、節の単位で編集、構造変換を可能にする。

表現の中心は意味レベルにおく。中心概念に対して、他の概念がある役割(格役割)をもって結合されている。例えば、「ヒロシはマリコに花束を贈った」という文は概念的には図2のように表現される。ここでは、「送る」という中心概念に対して、「ヒロシ」、「マリコ」、「花束」が、それぞれ、行為者(+主題)、受領者、対象という格役割をもってつながれていることが示される。また、これらの格は、この順序で「贈る」につながれていることを示す。また、それぞれをつなぐ助詞として、「は」、

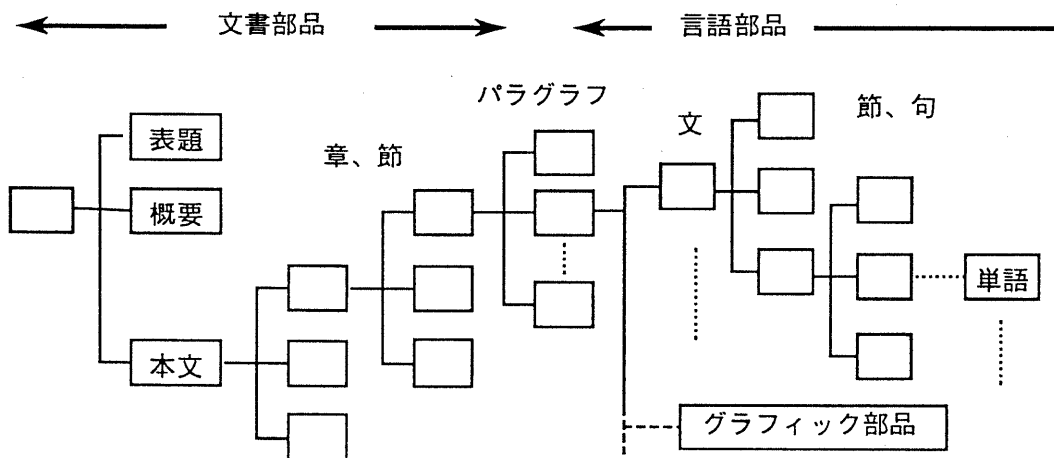


図1 文書部品と言語部品

「に」、「を」が用いられていることが示されている。

表層文を生成するために、構文情報、表層情報も必要になる。ここでは、部品間の結合に、助詞が記載されているのはこのためである。また、「ヒロシ」、「マリコ」、「花束」が、この順で「贈る」につながれていることも示されている。

この文に対して、「ヒロシ」を強調することを考える。強調には、「が」を使う方法と、「～したのはヒロシだ」のような強調構文を用いる

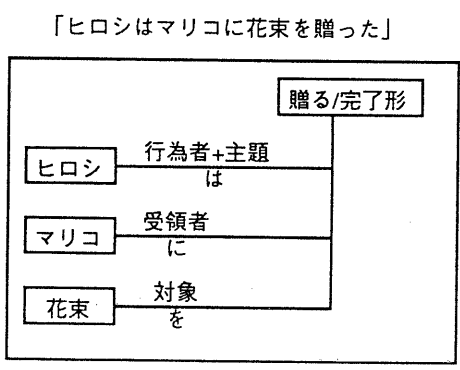


図2 言語部品表現の例

ことが考えられる。図3は、「ヒロシ」に対して、強調コマンドを発行し、後者の方法を選択したところである。

これは、効果を変えることを目的として、直接その目的をコマンドとして発行する例であった。一方、通常の文の編集は、その意図は頭の中であって、直接文に手を加えることになる。しかし、ここでは、予期しない効果が現れることもある。

例えば、ヒロシよりもマリコが社会的地位が高いことを示すために、「贈る」に対して謙譲語を用いることを考える。ここで、「贈る」の代わりに「献上する」を用いたとすると、システムは、「「献上する」は神仏や高貴な人に対する「やる」の謙譲語。大げさなイメージまたは皮肉的なイメージが生まれます。』のようなメッセージを出す。これが書き手が予期しなかった効果である。もちろんこの効果が意図にあっていれば、これ以上書き換える必要はない。例えば、「阪神はヤクルトにさらに2点を献上した」という文では、皮肉を込めることを意図しており、この場合では意図どおりということになる。

このように、X-ACTはテキストに対する編集操作に対して、その効果を指摘することを特徴とする。書き手は、システムからのメッセージを見て、テキストの適切さを判断しながら、テキストを表現意図に近いものに順次近づけていく。

3. 言語部品表現の設計

ここでは、言語部品に必要な特性を検討し、実現の方法を検討する。

推敲支援を考えた場合、言語部品が持つべき特性には、次のようなものがあると考えられる。

- 組替えが自由にできること
 - 種々の表現を試すことができる
- 書換えによる効果を知ることができる
 - 部品ごとのニュアンスの違いが記述され、比較できる
- 適切な文の構成、単語の使用が分かる
 - ただし、不適切な文も許す(不適切さが指摘されるのみ)

3.1 表現の枠組み

言語部品は、意味レベルの表現を基本にしている。しかし、構文の変化に対応することができること、表層文が見える形になっていることが必要となるため、意味、構文、表層の情報を含んだマルチレベル表現である必要がある。

例えば、HPSG [Pollard & Sag 1987]では、マルチレベル表現が可能になっている。しかし、表現がユニフィケーションで作られるため、作られる表現の中の要素は、もともと自分でもっていた(辞書から得られた)情報と、文の中の別の要素から得られた情報が融合されたものになる。このため構造の変更は容易ではないと考えられる。

言語部品表現は、構造の変更を重視するため、ユニフィケーションは用いない。部品には、それがつく部品に対する制約と、部品が結合したときに部品に与えられる操作が記述される。部品が結合されたり構造に変更が加えられるときに、それぞれの操作を起動することになる。

例えば、助動詞は、前に付く用言の活用の種類と活用形を規定する。受身の「れる」と「られる」の場合、前者は五段とサ変、後者はその他

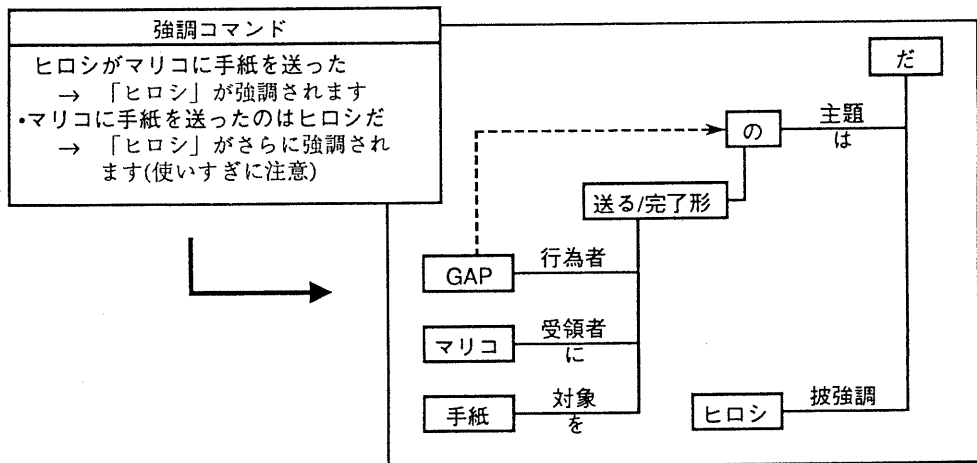


図3 言語部品による編集例

の未然形に付く。この助動詞を結合する場合、部品によって決まる不変の部分である活用の種類が検査され、可変の部分である活用形が決定される。

この可変部と不変部の違いを表すため、オブジェクト指向表現を用いる。不変部は、オブジェクトが共通に持つ(クラスがもつ)ものであり、クラスに与えられる。一方、可変部は各インスタンスがもつように規定する。

3.2 言語部品のもつべき情報

ここでは、推敲支援を行うために、言語部品がどのような情報を持たねばならないかを考察する。

言語部品表現のインスタンスからは、まず最低限文法的に適切な文が生成される必要がある。このため、形態素レベルでは、品詞、表記、活用の種類と現在の活用形が必要となる。構文レベルの情報では、自分自身に掛かる要素(例えば名詞+格助詞)の並びと、付けられている助動詞、助詞の並びが必要になる。

次に、適切な文を構成するための情報が記述されなければならない。形態素レベルでは、可能な表記の候補、およびその中で最も望ましい表記が示される。構文的には格パターンが必要である。これは、格の適切な順序、格助詞とその省略可能性、格役割、格に入る要素の意味的な選択制限が記載される必要がある。

さらに、適切な言葉、構文を選択し、書き手が表現したいと考えている文章を作ることをサポートできなければならない。例えば、書換えを行うために、部品間の関係の記述と、それぞれの意味やニュアンスの差の記述が必要になる。これらは、同義語、類義語(敬語も含む)や反義語などのシソーラスに、部品の要素の関連(格パターンなど)を加えたものになる。

3.3 言語部品の情報の保持方法

これらの情報は、オブジェクトの中のスロット値として保持される。先に述べたように不変部

はクラススロットに保持され、可変部はインスタンススロットに保持される。

また、一部のものは、クラス(タイプ)そのものがこの情報を担っている。例えば、動詞の活用の種類ごとに一つのクラスが定義され、各動詞はいずれかの活用の種類クラスを継承するようにする。活用の種類クラスには、語形変化を決定するメソッドが与えられている。

部品は、その機能に応じて、異なったクラスのサブクラスとなっている。

LP-CORE	メインの概念となりうるもの 動詞、名詞などの自立語 アクセスのための部品IDをもつ
LINK	部品の結合の手 格助詞等が対応する
LP-AUXV	助動詞、終助詞
LP-COMP	複合部品。以上のクラスのインスタンスの組み合わせ

これらのほかに、メインの概念にならないものとして、ギャップやゼロ代名詞のためのクラスが存在する。

3.4 言語部品表現の例

部品の定義例を図4に示す。定義中のシンボルはスーパークラス名、リストはスロット名とその値である。

ここでは、「勉強(する)」がサ変名詞であり、サ行変格活用を定義するクラス(Infl-Doshi-Suru)を継承していることを示している。格パターンは、

(品詞名 (助詞 省略可能性 格役割)...))

という形式の並びである。この部品が、ある品詞として働くときに、どういう助詞の並びを取るかを示し、そのときにその助詞でマークされた部品(名詞)が、どのような格役割でこの部品に繋がるかを示す。省略可能性は、この助詞が

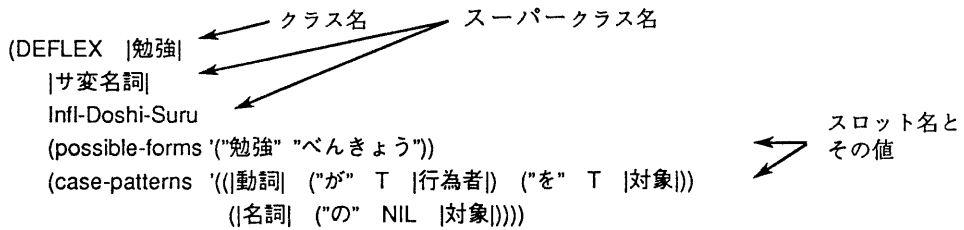


図4 言語部品定義例

省略可能か否かを示す。なお、ここでは、動詞、名詞それぞれに1つの格パターンの候補しか記述していないが、通常は複数個存在する。

4. 推敲支援機構

推敲のプロセスにおける書き換えは、

- 1) ユーザによるテキスト入力/変更指示
- 2) システムによる変更結果の候補の提示
および、それぞれの候補の評価
- 3) ユーザによる選択

の3つのステップから構成される。

このような目的を達成するため、個々のテキスト変更コマンドは、以下のステップがひとまとまりになったものを単位とする。

- 適用可能性評価
そのコマンドが実行できるかどうかを評価する関数
- テキスト変更方法
実際にテキストに仮の変更を加え、変更後の部品表現と、効果の評価を得る

一つのテキスト変更コマンドに対して、これらのメソッドは複数個存在することになる。テキスト変更コマンドが発行されたら、これらから複数の変更結果候補を得て、その評価とともに提示し、ユーザからの選択を受け付ける。最後に、この変更結果を、文書中で置き換える。

5. システムのUIと操作例

システムは、文書ウィンドウ、部品編集ウィンドウ、評価ウィンドウからなる。文書ウィンドウで選択された文(複数)が、部品編集ウィンドウに図示され、その評価が評価ウィンドウに示される。図5aにシステムの画面を示す(ここでは評価メッセージが出されていない)。

テキストに対する操作は、部品編集ウィンドウにおいて

- 部品を選択(複数)する
- コマンドの発行(メニュー)
- 変更結果から選択する

ことによってなされる。図5bに、結果の選択をしているところを示す。

6. 今後の課題

6.1 システムの拡充と評価

このシステムは、プロトタイプの開発途上であり、評価を行うまでに至っていない。今後、評価までに以下のことを行う必要がある。

•推敲支援機能の拡充

現在の提供している機能はプリミティブなもので、当初の目標とする「ユーザの意図を反映したテキスト表現の作成」は達成されていない。このためには推敲支援機能の拡充が必要である。

・辞書の拡充

通常、文を作成するのに困らない程度のエントリをもつ辞書を作成するだけでなく、未知語登録機構も必要なる。また、部品間関係を記述した辞書も必要になる。IPAL辞書、EDR電子化辞書の活用を考えたい。

・システムUI

現在、入力機構は非常にプリミティブなものしか用意していない。かな漢字変換によりテキスト入力を行うことと比較して、それほど負担が増えずに部品の組立てができるようになる必要がある。

また、部品の接続関係を明示的に刷るのではなく、通常テキストに近い形での表示編集も望まれる。

6.2 言語部品の他の応用の検討

言語部品は推敲支援を目指して設計しているが、以下の用途にも使える可能性がある。

・英文作成支援システム

言語部品を、英語の表層文を表現するものにする。個々の部品の意味、およびシステムからのメッセージを日本語で出す。これにより、英文の作成、書き換えを支援するシステムにすることができる。

・言語教育システム

上記システムを使っていくうちに英語の正しい使い方を学習することができる。すなわちこれは、英語の教育システムと見なすことができる。日本語部品に、外国語での意味/メッセージを組み合わせれば、外国人のための日本語教育システムとなる。

7. 他研究との関連

7.1 推敲支援

推敲校正支援システムでは、字面レベルでは「推敲」[牛島1991]がある。これは、構文解析などを行っていないので、誤りの可能性のある

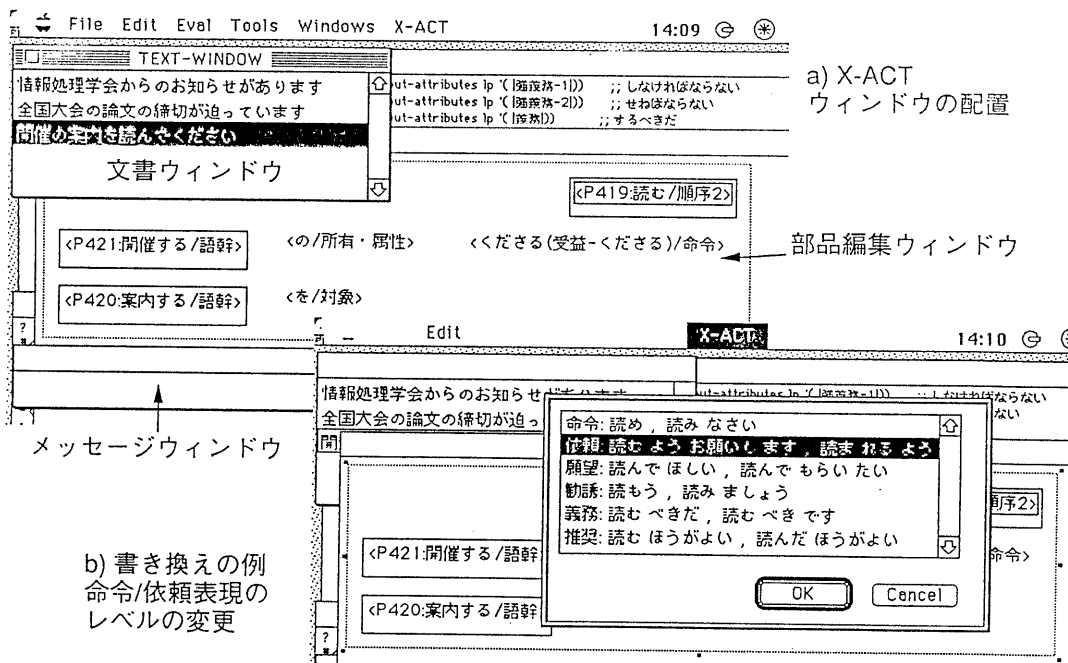


図5 システム画面

ものをすべて表示する。例えば、あいまいな接続助詞「が」の指摘をする機能では、すべての「が」が列挙される。推敲のプロセスのすべてにおいて、結局自分で判断することになる。

形態素レベルの解析を行うものとしてCOMET [福島他1986], CRITAC [鈴木他1986], FLeCS [奥村他1992a, b] などがある。推敲支援規則を形態素解析した結果に対するパターンとして記述する。これにより、誤りだけでなく、不適切な表現である可能性を指摘する。しかし、書き手の表現意図にあった適切な文を作成するという視点はない。この点は、構文解析までおこなうMAPLE [高橋他1990]も同様である。

書換えを支援するものとしては、REVISE-S [林他1991, 武石他1992]がある。構文レベルの解析を行い、不適切さを指摘し、必要に応じて書換えを行う。このため、文の分割や、二重否定など、高度な意味判断を要するものも可能になっている。しかし、これも書き手の表現意図にあった文を作成するという視点はない。

7.2 テキストの構造化

テキストの構造化に関しては、FleCS [奥村他1992], COMET [福島他1986]で対応している。これらは、形態素の並びとして表現されており、推敲すべき場所の検出には効果があっても、文の書換え、構造変換には適切ではない。

テキストの構造化としては、むしろコーパスにおける例が、構文にまで対応している点に近い。これはATR [江原他1992], BSO [Sadler 1991]などで行われている。しかし、これらも、目指すところは言語データとして利用しやすくすることであり、我々の目的とは異なる。

8. おわりに

ここでは、推敲とは、テキストに対して書換えを行い、その効果を見ながら最適な表現を選択するプロセスと位置づけた。この推敲のプロセスの支援を行うための枠組みとして、言語部品表現を提案した。言語部品表現により、文の句

や節、単語単位での編集を容易にすることができる。

先に述べたように、今後は、この枠組みを用いて、推敲支援機能の充実を図る必要がある。また、部品の充実、部品間関係辞書など、部品に関する知識の充実が必要である。

最後に、有益なアドバイスをいただいた篠岡主幹研究員、上林主幹研究員を初めとするシステムコミュニケーション研究所の諸氏に感謝します。

参考文献

- 牛島:「日本語文章推敲支援ツール「推敲」」bit Vol.23, No.1 (1991).
- 江原、小倉、篠崎、森もと、樽松:「電話またはキーボードを介した対話に基づく対話データベースADDの構築」情報処理学会論文誌 Vol.33, No.4 (1992).
- 奥村、建石、脇田、金子:「日本語校正支援システムFleCSの新聞社における実用化」情報処理学会自然言語処理研究会87-11 (1992).
- 奥村、脇田、金子:「日本語校正支援システムFleCSの新聞社における実用化」情報処理学会自然言語処理研究会91-5 (1992).
- 鈴木、武田、藤崎:「日本語文書校正支援システムCRITAC」情報処理学会日本語文書処理研究会8-5(1986)
- 高橋、吉田:「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Vol.31, No.7 (1990).
- 武石、林:「接続構造解析に基づく日本語複文の分割」情報処理学会論文誌 Vol.33 No.5 (1992).
- 林、菊井:「日本文推敲支援システムにおける書き換え支援機能」情報処理学会論文誌 Vol.32, No.8 (1991).
- 福島、大竹、大山、首藤:「日本語文章作成支援システムCOMET」信学技報OS-86-21 (1986).
- Sadler, V.: "The Textual Knowledge Bank: Design, Construction, Applications". in Proc. Int'l Workshop on Fundamental Research for the Future Generation of Natural Language Processing, Kyoto (1991).
- Pollard, C. & Sag, I. A.: "Information-based Syntax and Semantics, Volume I Fundamentals," CSLI Lecture Notes Number 13, CSLI (1987).