

## 日本語を対象とした 文評価システムに関する研究

箱守 聡      杉江 昇      大西 昇  
(名古屋大学工学部)

事実の伝達を目的とする文章では、正確さ、わかりやすさが要求される。人間がこれらの文章を作成するときには、一度書いた文章を読み直し、修正する、添削作業を行なうのが普通である。本研究では、この添削作業を支援するシステムについて考察する。規則の根拠として、従来用いられてきた文章作成の指導書を参考にした。さらに、1文を対象として、統語レベル、意味レベルの処理を行ない、修飾関係のあいまいさ、わかりにくさを指摘するシステムを試作した。

## A Correction System for Japanese Text

Satoshi HAKOMORI Noboru OONISHI Noboru SUGIE

Nagoya University  
Furo-cho, Chikusa-ku, Nagoya, 464  
JAPAN

During the preparation of manuscripts such as research articles, business letters, etc., the detection and correction of imperfect sentences take place frequently. In this paper, we introduce a system which can detect and correct certain imperfect Japanese sentences.

The system has 2 functions. First, it detects the syntactic ambiguities in the given sentences. After syntactic analysis, semantic analysis is carried out to reduce the ambiguity. If more than one structures remain, then the system points out that the sentence is syntactic ambiguous. Second, the system detects less readable orders in consecutive modifiers and indicate more readable ones. This process is based on several rules. We show some examples demonstrating the ability of the system.

## I. はじめに

我々が日頃よく目にする、新聞記事、評論、論文、解説などのいわゆる実用的文章は、書き手より読み手へ、ある事実関係や書き手の考えを伝えるために存在するものである。したがって、これらの文章は、何よりも読み手にとって読みやすく、わかりやすいものでなければならない。また、書き手が伝えなかった事柄が、読み手に正確に伝わらなければ意味がない。

文章を作成する過程には、通常、1度作成された文章を読み直して修正する、添削あるいは校閲という作業がある。この作業は、正確な、わかりやすい文章を作成する上で重要なものであると考えられる。

添削作業において、修正を必要とするような文を、本論文では悪文と呼ぶことにする。悪文は、言語現象のあらゆるレベルにわたっていて、その修正を行なう際も、複数のレベルの処理を要することが多い。

本研究では、この添削の作業に注目し、与えられた日本語文に対してそのあいまいさ、わかりにくさの評価を行ない、文章作成支援システムについて考察を行なった。このような機能を持った英語・日本語ワードプロセッサは、最近実用化され始めている[1][2][3][4][5]。しかし、それらによって指摘される現象の範囲は、おもに語彙処理の段階となっている。ここでは、統語処理及び意味処理レベルの解析を行ない、その範囲でいくつかの悪文の現象が指摘され、その一部はよりよい文に修正されることを示す。

## II. 悪文とは

### 2. 1 文章作成の流れ

人間が文章を作成するとき、おおよそ図1に示すような流れに沿って作業を行なっていくものと考えられる。このうち、主題の決定、構成の決定までは、頭の中あるいはメモなどで行なわれ、文章作成に先立つ作業である。ここでどの様な文章を書くか、大まかな構成が定まると、まず実際に下書きを書くであろう。そして、書いてみた文章が自分の意図した

内容を十分表わしているかどうか、作者自身あるいは他の人によって検討され、必要な処理がなされる。これを添削過程と呼ぶ。

この添削は、大きく2つの作業に分けられる。

① 修正を必要とする文、あるいは箇所を指摘

② ①の指摘に基づいた、より良い文章の作成  
添削の作業は、文章作成の諸作業のうちでも、最も客観性を必要とするものであると思われる。文によって表現される内容は、作者には周知のものであるため、文章に欠陥があったとしても、書いた本人がそれを指摘することは難しい。添削を行なうときには、作者以外の誰かが文章を読み、修正箇所を指示するほうがより良い結果を得ることができるであろう。したがってこの作業を自動化することは、文章作成を計算機により支援する上で効果的であると考えられる。

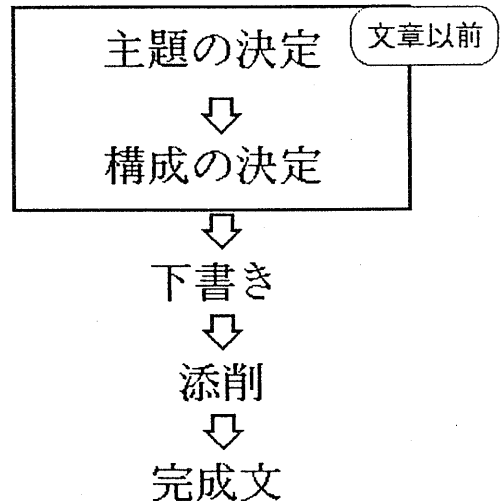


図1 文章作成の流れ

## 2.2 悪文の分類

実際に文章の添削を行なうには、まず修正を必要とする文、すなわち悪文を捜し出さなければならぬ。

自然言語処理の観点からみれば、語彙、統語、意味、談話の各解析レベルによる分類が考えられる。しかし、後に示すように悪文の現象は、その指摘、修正どちらにおいても複数のレベルの処理を必要とするものが多く、悪文をこの基準で分類するのは適当でない。

悪文の評価、分類については、今までに、言語学、日本語学等の方面で研究がなされている。また、文章の書き方に関する実用書も多く出版されており、その中である程度の分類がなされている。

岩淵ら[6]は、「悪文をさけるための五十カ条」として、悪文を作らないための心得を大きく4つに分類している。筆者らはまず、この分類をもとにし、さらに[7][8]などによる評価を含めて、悪文の具体例の整理を試みた。これを図2に示す。ただし、岩淵らが「敬語の使い方に関するもの」とした部分は、「文章構造による悪文」の一例として扱っている。

次に、文章を読んだときの理解のしかたにより、悪文を次の3つに分類してみた。

- ① その文が理解できない、または矛盾を持つと

## 感じる文章

- ② 一読して内容が理解できない、または理解に必要な以上にかかる文章

- ③ 理解できる内容が複数できる文章

①に当てはまる文章は、その構造になんらかの誤りがある場合である。この誤りが談話・意味レベルで起こったときは、一つ一つの文の内容は把握できても、文章全体では矛盾を生じる。また、統語レベルで起こるときは、その文が文法的な間違いを犯して、文として機能していないときであると考えられる。このような文は非文と呼ばれ、通常は文とはみなされない。非文を指摘すること自体は比較的簡単であるが、誤りの内容や修正の方法も提示するには、従来の文法規則だけでは困難である[3][4]。

②に当てはまる文章は、その内容を一つに特定することができるが、理解するまでに何度も読み返したり、頭の中で整理するなどの操作を必要とする場合である。このような文章では、同じ内容に対して、もっと他の良い表現が存在すると思われる。ここに当てはまるような文章は、本研究では「わかりにくさ」を持つという。例えば次の文は、統語的なわかりにくさを持つといえる。

分類	具体例
語の使い方による悪文	<ul style="list-style-type: none"> <li>言葉の誤用</li> <li>かたすぎる言葉の多用</li> <li>意味の重複した表現</li> <li>漢字と平仮名の割合が不適当</li> <li>あいまいな用語（「……的」、「……性」、など）の多用</li> <li>同一表現の多用</li> </ul>
文構造による悪文	<ul style="list-style-type: none"> <li>長すぎる文、修飾語</li> <li>修飾関係が不適当</li> <li>文のねじれ（副詞の呼応、助詞の取り扱いなど）</li> <li>句読点の位置</li> <li>複文における、話題の食い違い</li> <li>その他統語的誤り</li> </ul>
文章構造による悪文	<ul style="list-style-type: none"> <li>文脈の食い違い、矛盾</li> <li>文体の不統一（敬語、ていねい調、である調など）</li> <li>接続詞の間違い、不適当な使用</li> </ul>

図2 悪文の具体例

私は太郎が次郎が三郎が死んだ現場にいたと証言したのかと思った。(1)

③に当てはまる文章では、文法的、意味的、談話的にみても特に問題ないと思われる解釈が複数できる。悪文でのこのような現象を、本研究では「あいまいさ」を持つと言うことにする。例えば、次の文章は文の意味レベルにおいてあいまいさを持つと考えられる。

刑事は泥まみれになって逃げる犯人を追いかけた。(2)

(2)では、泥まみれになったのは刑事なのか、犯人なのかは、この文からは決定することはできない。しかし、この前後に他の文章があったとき、その内容からこの文の内容を一つに決められることもある。そのようなときは、談話レベルにおいてはあいまいさは発生していない。

### Ⅲ. 文評価システム

#### 3. 1 試作システムの対象

以上述べてきた悪文に対する添削を実際に行なうようなシステムを本研究では試作した。ただし、すべての現象を対象として評価を行なうシステムの実現は難しいので、次のように対象の制限を設けた。

入力する文章の対象として、

- ① 埋め込み文の存在を許す、一つの文であること。ただし、接続助詞のある文は除く。
- ② 文法的に正しい文であること(通常の文法により構文解析が行える文であること)。

このような入力文に対する添削処理として、

- (1) 修飾関係におけるあいまいさの検出
- (2) 修飾関係における、語順によるわかりにくさの検出、修正

入力文の対象を上のように制限することによって、文の解析は統語レベル、意味レベルの問題となる。すなわち、本システムでは、談話レベルの処理は行っていない。

次に、上に示した添削処理について、詳しく述べる。修飾関係は、連体修飾、連用修飾の2つに大きく分けることができ、それぞれによって処理法が若干異なる。

#### 3. 2 修飾関係におけるあいまいさの検出

連用修飾、連体修飾共に、修飾句が、統語的にも意味的にも係ることのできる被修飾語を複数持つときに、あいまいさが起こる。このとき、統語的に係り得ることを示した文の解析木に対し、意味的にも係り得るかどうかの判定を行い、あいまいさの現象をはからなければならない。

連用修飾の場合、被修飾句は述語であり、修飾句は述語に対してなんらかの格を持つと考えられるから、述語についての格フレームを用意し、各々の修飾句がその格スロットに矛盾なく埋め込まれるかどうかで、その意味整合性を判定することができる。しかし接続助詞によって修飾文が述語にかかっているような場合には、修飾文の述語と被修飾語の述語との意味整合性を調べなければならず、談話的な知識を必要とする。例えば、

大声をあげて騒ぐ子供をたしなめた。(3)

大声をあげて寝ている子供を起こした。(4)

という2つの文では、文の形は同じであるが、(3)ではあいまいさが起こり、(4)では起こらない。これは、(3)では「大声をあげる」と、「騒ぐ」、「たしなめる」との両方が意味整合性を持つが、(4)では「大声をあげる」と「寝る」との意味整合性がないためだと考えられる。本システムでは、このような述語同士の意味整合性を調べることはしていない。したがって(4)でもあいまいさを持つ判定をする。

連体修飾の場合は、被修飾句は名詞句である。このときは、修飾句と被修飾句との意味整合性を調べることが必要である。平井ら[9]によれば、この意味関係は大きく5つに分類ができるとしている。本システムは、このうち「格要素型」の意味関係を持つ連体修飾の解析を行なう。

連体修飾のあいまいさが起こる形に、「修飾句+N<sub>1</sub>のN<sub>2</sub>」という名詞句がよくあげられる。修飾句が、N<sub>1</sub>、N<sub>2</sub>のどちらにもかかることができるとき、あいまいさが起こる。この場合の例として、

きのう家に来た太郎の妹はかわいい。 (5)

という文を考えてみる。(5)の文だけでは、きのう家に来たのは、太郎なのか、太郎の妹なのかを決めることはできず、あいまいさを持つことになる。この解析は、被修飾句のそれぞれが修飾句中の述語「来る」の格要素となることができるかどうかを調べることによって行なう。

この方法を用いたとき問題となるのは、例えば「大きい鼻の男」という名詞句についても、あいまいさがあると判定してしまうことである。この場合、人間が読むと、「大きい」は「鼻」にしかかからず、あいまいさは起こらない。本システムでは、このようなときに誤った判断をしないために、「N<sub>1</sub>のN<sub>2</sub>」の名詞句において、N<sub>1</sub>がN<sub>2</sub>の部分関係を持ち、かつN<sub>1</sub>が限定された概念を持っていないときに、その名詞句は意味的に存在できないとしている。

### 3.3 修飾関係におけるわかりにくさの検出

連体、連用ともに、修飾関係によるわかりにくさは、その文の語順によるところが大きいと思われる。そこで本研究では、わかりにくさを語順に限って扱うことにした。

[6][7][8] などによれば、わかりやすい文章を書くとき、語順について次の点を守るようにするとよい。

- ① 修飾語のすぐ後に被修飾語をおくこと。
- ② (同じ語にかかる修飾語が複数あるとき) 長い修飾語を前に、短い修飾語を後にすること。
- ③ (同じく修飾語が複数あるとき) 大きい状況を示す語を前に、小さい状況を示す語を後にすること。

規則の適用は、原則的に、①②③の順に優先度が高いとする。また、②の長い、短いという基準は、文節の数を比べることとする。

③の大状況、小状況については、本研究では、被修飾句に対し、その性質、属性を限定するするような修飾句を小状況とし、具体的な分類として、形容詞、形容動詞を述語として持つ修飾句をその範囲と規定した。

では、実際にこの規則を用いて文を評価してみる。①を破る文は、主に前出の(1)のような、いわゆる入れ子構造となっているものである。この文の係り受けを図3に示す。(1)は、次のように語順を入れ換えると読みやすくなる。

三郎が死んだ現場に次郎がいたと太郎が証言したのかと私は思った。 (1)

これは、①にしたがい、連用修飾句を述語のすぐ近くに置くようにした文である。次に、

太郎が私の妹に次郎がきのう山で見つけた花をあげた。 (6)

この文では、「あげた」という述語に3つの連用修飾句がかかっている。これらを、②にしたがって文節数の多い順に並べかえてみる。

次郎がきのう山で見つけた花を私の妹に太郎があげた。 (6)

1文でみる限り、こちらの方が読みやすくなっていると思われる。最後に、③の適用の例として、

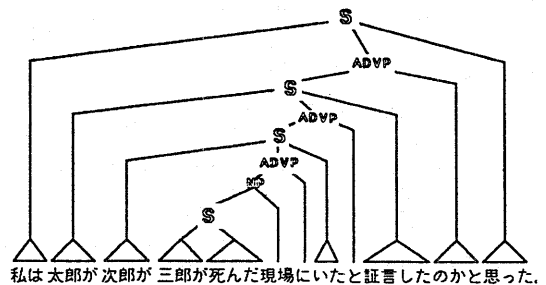


図3 (1)の係り受け

白い香りの良い垣根に咲く花を見つけた。(7)

という文を考えてみる。「花」に対し連体修飾句が3つかかっている。「白い」は他の2つに比べ短いので、後に置く。「香りの良い」と「垣根に咲く」はともに同じ長さだが、前者は花の属性を示しているのに対し、後者はその場所を示している。したがって前者の方が花に対しより小さい状況を表わしていると考えられるので、こちらを後ろに置く。書き換えた文は、

垣根に咲く香りの良い白い花を見つけた。(7)

この方がより読みやすいと思われる。

### 3.4 試作システムの構成

本研究で試作した文評価システムの構成を図4に示す。環境上の制約のため、入力文は現在ローマ字で書かれている。文はまず統語解析され、その結果が意味解析部に渡される。意味解析によって抽出された意味構造は、文評価部で評価される。統語解析、

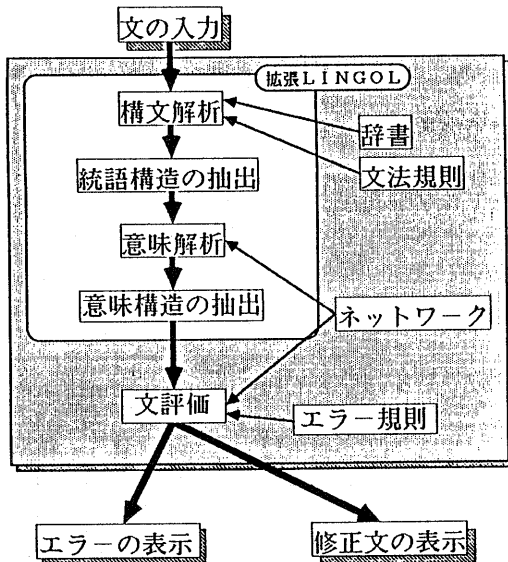
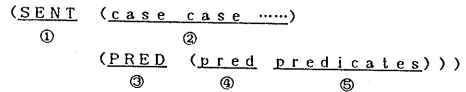
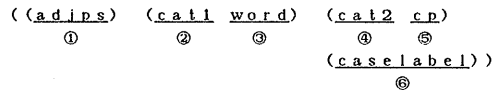


図4 システムの構成



- ① このリストが文であることを示すラベル
- ② 文の述語のかかる修飾句のリスト。図5、図6のどちらの構造も要素となる。
- ③ 述語のリストであることを示すラベル
- ④ 述語中、意味を支配する語(用言)の終止形
- ⑤ 文に現われる述語の文字列

図5 文を表わす意味構造



- ① 連体修飾句のリスト。図5、図6のどちらの構造も要素となる。
- ②, ③ 被修飾語。②はnoun, np, advのどれかで、③がその語。npとなるのは並列句のとき。
- ④, ⑤ この句につく格助詞。④はそのカテゴリー、⑤はその語。
- ⑥ この句が、連用修飾句のときに持つ格ラベル

図6 修飾句を表わす意味構造

意味解析には、電総研で開発された拡張LINGOLを用いた[10]。次に、各部の詳細について述べる。

#### ◎ 構文解析部

辞書、文法を用いて文の統語構造を明らかにする。ここでは、入力文について統語的に正しい構造は全て得られるようになっている。

#### ◎ 意味解析部

構文解析部では、解析木が一般に複数生成される。ここでは、それぞれについて修飾関係の意味整合性を調べながら、その意味構造を作成する。意味整合性を調べる際には、述語の格フレームの他に、名詞の上位概念を探ることが必要になる。本システムでは名詞間の意味ネットワークを作成して行なっている。

本システムに用いた意味構造は、図5、図6の2つの構造の再帰的な組合せにより成る。図5は文構造を、図6は連体・連用修飾句構造を表わす。このような構造にするには、次の点を考慮した。

- ① 一つの語にかかる修飾句をひとまとめに取り扱えること。
- ② 入力文の語順を復元できること。

この意味処理の過程で、意味的に存在しえない解釈を削除するようにした。こうして意味解析が終了した時点で、意味構造が複数存在するならば、入力文にあいまいさが発生しているとみなす。

◎ 文評価部

文評価部では、まず意味解析部より送られてきた意味構造が1つであるかどうか調べる。複数ある時は、あいまいさがあるとして、エラーを表示する。これらのうちいづれかが入力者の意図した意味となるはずであるが、どれであるかをシステムが決定することはできない。本システムでは、現在は得られた意味構造をそのまま表示し、入力者に選択させるようになっている。

1つに絞られた意味構造は、次にわかりにくさの評価を受ける。ここでは、3.3 に述べられた規則を用いるが、連体修飾では①②③、連用修飾では①②

を適用する。規則に合わない表現であると判定したときは、規則に合わせ、語順を入れ換えた構造を新たに作り、それより表層文を再構成して表示する。

IV. 結果

本システムを用いて、実際に文の評価を行なった例を図7、図8に示す。

図7は、(5)の文を入力したときである(①)。システムは2つのあいまいさを見つけ(②)、その意味構造を番号をつけて表示する(③)。入力者がそのうち、1を自分の意図した意味として選択すると(④)、その構造ではわかりにくさは起こっていないことを知らせる(⑤)。

```

①  s_
    KINOU IE NI KITA TAROU NO IMOUTO HA KAWAII.
    *****
②  THIS SENTENCE HAS 2 AMBIGUITIES.
    1 (SENT (((SENT ((NIL (ADV KINOU) NIL (TIM)) (NIL (NOUN IE) (CP12 NI) (STO)))
          (PRED (KURU KITA)))
        (NIL (NOUN TAROU) (CP11 NO)))
        (NOUN IMOUTO)
        (CP12 HA)
        (SUB)))
      (PRED (KAWAII KAWAII)))
③  2 (SENT ((((((SENT ((NIL (ADV KINOU) NIL (TIM))
          (NIL (NOUN IE) (CP12 NI) (STO)))
          (PRED (KURU KITA)))
        (NOUN TAROU)
        (CP11 NO)))
        (NOUN IMOUTO)
        (CP12 HA)
        (SUB)))
      (PRED (KAWAII KAWAII)))
    CHOOSE YOUR MEANING NUMBER ?
④  s_
    1
    <WAKARINIKUSA> NOT OCCURS.
    KINOU IE NI KITA TAROU NO IMOUTO HA KAWAII
    -----CRITIQUE END
  
```

図7 (5)の文の添削結果

```

①  s_
    TAROU GA WATASHI NO IMOUTO NI JIROU GA KINOU YAMA DE MITSUKETA HANA WO AGETA.
    *****
②  THIS SENTENCE HAS NO AMBIGUITY
③  <WAKARINIKUSA> OCCURS IN THIS SENTENCE.
    BETTER STRUCTURE.....
④  JIROU GA KINOU YAMA DE MITSUKETA HANA WO WATASHI NO IMOUTO NI TAROU GA AGETA
    TAROU GA WATASHI NO IMOUTO NI JIROU GA KINOU YAMA DE MITSUKETA HANA WO AGETA
    -----CRITIQUE END
  
```

図8 (6)の文の添削結果

図8は、(6)の文を入力したときである(①)。この文には、あいまいさはないことを知らせ(②)、次にわかりにくさがあることを知らせる(③)。そして、規則にしたがって語順を入れ換えた、よりわかりやすい文を提示し(④)、終了する。

## V. 考察及び課題

本研究で試作したシステムで、3.1 で定めた対象については、ほぼ満足のゆく結果を得ることができた。

本研究の課題を挙げる。

あいまいさを見つける方法として、本研究では、統語的に可能な解析結果を全て出し、その一つ一つが意味的に正しいかどうかを調べている。この方法では、無駄な解析に時間をとられる恐れがある。統語解析と意味解析を一緒に行なうようにする方がより早い処理となる。しかし、その場合には、必要な結果を解析の途中で削除してしまわないよう、工夫しなければならない。

悪文は、複数の解析レベルにわたって添削すべきものが多い。本研究では、統語・意味処理に限った添削を行なったが、あいまいさ、わかりにくさ共に、完全な添削をするには、談話的処理が必要となる。例えば、(5)の文において、その文が現われる以前に太郎と妹のどちらかがその家に行っていることが記述されていれば、あいまいさは起こらなくなる。また、(6)の文でも、談話の流れとして、太郎のことが主題となっていれば、「太郎が」という句を最初に持って行く方が自然であろう。したがって実際には、本研究で示した規則と、談話的な添削規則とをうまく用いて行かなければならない。

また、本研究では修飾関係に限った添削を行なったが、この場合、句読点を考慮する必要もある。読点の使用により修飾先を決定したり、わかりにくさを避けることも可能である。

## 参考文献

- [1] Macdonald, N.H.: "The UNIX Writer's Workbench Software: Rationale And Design", bell Syst. Tech. J., Vol.62, No.6, pp.1891-1908 (1983)
- [2] Heidorn, G.E., Jensen, K., Miller, L.A., Byrd, R.J. and Chodrow, M.S.: "The EPISTLE text-critiquing system", IBM Syst. J., Vol.21, No.3, pp.305-327 (1982)
- [3] 河合, 杉原, 杉江: "英文の誤りを検出するシステムASPEC-I", 情処学論, Vol.25, No.6, pp.1072-1079 (1984)
- [4] 河合, 杉原, 杉江: "英文添削のための対話型システムASPEC-II", 情処学論, Vol.26 No.5, pp.815-823 (1985)
- [5] 日経エレクトロニクス: "添削支援機能を備えたワードプロセサ、日立製作処とリコーが発売", No.436, pp.124-125 (1987)
- [6] 岩淵悦太郎編: "第三版 悪文", 日本評論社, 1977
- [7] 本多勝一: "日本語の作文技術", 朝日新聞社, 1976
- [8] 相原林司: "文章表現の基礎的研究", 明治書院, 1984
- [9] 平井誠, 北橋忠広: "日本語文における「の」と連体修飾の分類と解析", NL-58-1, 1986
- [10] 電総研推論機構研究室: "拡張LINGOL", 1978