

## 科学技術系論文における付属語連鎖の 統語的、意味的な誤りの検出方法

蓮井 洋志<sup>†</sup> 川口 湊<sup>††</sup> 小倉 久和<sup>††</sup>

日本語の中で複雑で誤りが含まれやすい表現として付属語連鎖がある。我々は科学技術系の論文の推敲支援システムの一環として、付属語連鎖の正誤を判断するパーザを開発した。このパーザは、本稿で述べる統語規則、意味規則をもとに構成されている。統語規則は既存の付属語の統語構造についての佐伯の解析結果をもとに作成したが、科学技術系論文にふさわしいと思われる付属語連鎖を受容するように改良した。意味規則は付属語に意味属性を与え、意味属性の矛盾、重複、呼応関係の誤り、語用法の誤りを表現したものである。この付属語パーザは、統語規則の受容しない付属語連鎖と意味規則によって誤りを含むと判断された付属語連鎖を推敲の必要な表現として検出する。この論文推敲支援システムを国語学の論文および口語体の文章に使用した結果、口語体の文章または国語学の論文には、科学技術系の論文には見られない表現が多いため、本システムで多数の付属語が誤りとして検出された。

### Error Detection Method by Use of Syntactic and Semantic Rules for Chains of Agglutinated Words in Japanese Scientific Papers

HIROSHI HASUI,<sup>†</sup> MINATO KAWAGUTI<sup>††</sup> and HISAKAZU OGURA<sup>††</sup>

Chains of agglutinated words in Japanese are complex and difficult expressions. We implemented a computer-aided refinement system for scientific articles. The system has Agglutinated Word Parser that judges whether a string is identified with chains of agglutinated words in scientific papers. Grammar rules of Agglutinated Word Parser consist of syntactic rules and a semantic rules for agglutinated words. The syntactic rules of the parser are based on the result of the Saeki's analysis in his linguistic paper which proposes syntactic rules. The semantic rules reveal inconsistency, wordiness, mistaken agreement, and misuse in chain of agglutinated words. In this system, the chains which the syntactic rules do not accept or for which the semantic rules detect anomalies, need further refinement. The results of application of Agglutinated Word Parser to 13 papers show that some chains in ordinary writings do not exist in scientific writings.

#### 1. はじめに

日本語の付属語の接続関係は非常に複雑である。日本語を日常的に使用している我々でも、ときには誤った使い方をする。普段の生活での会話は曖昧な表現に満ちている。それは、日本人自体が曖昧な表現を好むという社会的な背景もあると思われる。しかし、科学技術系の論文は意味の曖昧性をなくすために、簡潔な付属語連鎖で表現することが必要である。

留学生の日本語初学者は、日本語の付属語連鎖の複雑さに苦しめられる。留学生の文章の中には、よく格助詞の使い方などを間違えていたり、必要な部分に付属語がないことなどがある。学生が作成した論文などは、誤字、脱字を含んでいるだけでなく、仮名漢字変換の誤変換なども含まれている。そういった文章は、付属語の表現にまで気を配られているかどうかは疑問である。研究者の論文においても、しばしば同様なことがある。また、翻訳された文章の一部が非常に読みにくい理由のひとつとして、付属語の使い方の問題がある。

付属語の統語規則はさまざまな場所で発表されている<sup>1)~2)</sup>。文献<sup>3)</sup>にはコンパイラ・コンパイラ yacc を用いた付属語の解析方法について書かれている。文献

<sup>†</sup> 福井大学工学部大学院工学研究科

Graduate School of Engineering, Fukui University

<sup>††</sup> 福井大学工学部情報工学科

Department of Information Science, Faculty of Engineering, Fukui University

2)には機械翻訳における付属語の意味属性や付属語の接続規則について分類してある。また、付属語の品詞の接続表なども、付属語の統語規則を表しているといえる<sup>4)</sup>。しかし、科学技術系の論文としてふさわしい付属語だけを受容することを目的とする付属語のパーザを作成した試みはない。

我々は論文推敲支援システム nspell を開発しているが、今回その一環として科学技術系の論文としてふさわしい付属語連鎖だけを受容することを目的とした付属語パーザを開発した。この付属語パーザは、次章で述べる統語規則と意味規則を実現した。付属語パーザを用いて形態素解析を行い、その結果を利用して、読点の打ち方、再考を要する表現の検出、誤字・脱字の検出、複合語の誤りの検出、仮名遣いの誤りの検出などを行う。現段階では構文的な誤りのチェックは行っていない。

本研究の付属語パーザは形態素解析の一連のながれに組み込まれている。あらかじめ付属語パーザには仮名遣いの誤りの規則や意味の矛盾の規則を用意しておく、形態素解析に失敗した部分や誤りの規則に合致した文字列を誤りとする。形態素解析情報や誤りの規則を利用している点では FleCS<sup>5)</sup>に近いが、検出対象が付属語連鎖である点で異なっている。付属語連鎖の誤りの検出方法としては、ほかに2重マルコフモデルを利用した方法<sup>6)</sup>や“推敲”<sup>7)</sup>などで用いられている字面解析の手法がある。2重マルコフモデルを用いた誤りの検出・訂正は形態素解析を行わずに文字列のマルコフモデルを用いて誤りを推定する。“推敲”では推敲の必要性の高い表現を文法解析によらずに抽出を行う。誤りを推定する推敲支援システムとしては他に MAPLE<sup>8)</sup>や CRITAC<sup>9)</sup>などがある。MAPLEでは自動翻訳システム ATLAS を利用して文章を文法的に解析し、用語の適切さのチェック、文の分かりやすさの評価、索引項目の抽出などを行う。CRITACでは統計的な手法を用いた複合語の仮名漢字変換の変換誤りの検出・訂正、文書スタイルの診断、校正の支援などを行う。

付属語パーザの統語規則は、文献1)にある付属語の接続関係を表した「文末の構成図」をもとに、統語規則を追加改良した。本論文で検討した意味規則は、主として不適格な意味上の構成をチェックする規則である。たとえば、1つの付属語連鎖内にある、2つの付属語の意味属性の重複、矛盾、呼応の破れなどを検出する。この意味規則で検出された付属語連鎖は、推敲の余地があると見なす。付属語パーザは、Freeware の Bison を用いて作成した。

本論文では、2章で付属語連鎖の統語規則と意味規則についてまとめる。3章は推敲支援システム nspell と付属語パーザについて、4章は実際の文章に対して付属語パーザを適用した評価結果についてまとめた。5章は考察と課題にあてた。6章はまとめである。

## 2. 付属語連鎖の統語規則と意味規則

本論文で検討する付属語パーザは文節単位で処理する。とりわけ述語を含んだ文節は、文の中で大きな役割を果たしているために、付属語連鎖の使い方で曖昧な表現になったり、意味が通じない表現になったりすることがしばしばある。科学技術系の論文には不似合いな付属語を含んだ付属語連鎖や誤字・脱字についても1文節の解析だけでかなり検出できる。

本論文では、自立語連鎖に付属語連鎖が接続しているものを文節という。そのとき、句読点およびかぎ括弧などの記号も付属語と考えて扱う。付属語が接続しない連体詞や副詞などの自立語はそれだけで1文節と考える。

### 2.1 付属語連鎖の誤りの例

付属語連鎖の誤りには以下のような例がある。

- (i) 工作するです。
- (ii) 兄たりがたく、
- (iii) 理解してと(頼んだ。)
- (iv) 移動するかもしれなさそうだ。
- (v) 開発してしまってきた。
- (vi) 大切であるそうではない。
- (vii) 問題ではある。

(i)は付属語の接続が日本語として誤りである。(ii)は文語的な表現である。(iii)は口語的な省略表現である。(iv)は表現が冗長である。(v)は意味が1文節内で矛盾している。(vi)は語の使い方が不自然である。(vii)は誤りではないが「問題ではあるが、」とか「問題である」という使い方をするのが普通である。つまり、提示助詞「は」と逆接の助動詞「が」が呼応するのが通常であるのに、この例では呼応していない。

これらは、科学技術系の論文の付属語連鎖としては不適當である。

### 2.2 付属語連鎖の統語規則

統語規則は付属語の接続関係を生成規則として表す。たとえば、「分からなかった。」は動詞「分かる」の未然形に否定の助動詞「ない」の連用形を接続し、その後過去形の助動詞「た」が接続した表現である。これを統語規則として、

- (a) (動詞) → 動詞の未然形, (「ない」: 否定)
- (b) (「ない」: 否定) → 「ない」の連用形, (「た」:

過去)

(c) (「た」: 過去) → 「た」の終止形, (句点)

(d) (句点) → 句点

のように表す. この例では, ( ) で括ったものは非終端記号である. 非終端記号の(動詞)の規則(a)は左辺の(動詞)を右辺で書き換える生成規則である.(b), (c), (d)についても同様で, 本論文で扱う統語規則は, すべてこのような正規文法の形式の生成規則となっている. 非終端記号内の“:”に続くものは意味属性を, “,”は要素の区切り目を表す.

文献10)を参考に, 統語規則で用いる付属語の接続上の分類を付録Aにまとめた. また, 用意した統語規則は1266個にのぼるが, その一部を付録Bに示す.

統語規則の例外として, 特殊な活用をする用言のための規則がある. 「ない」は形容詞で「そうだ」[様態]に接続して, 「なさそうだ」になるが, 普通の形容詞は, 「遠さそうだ」とはいわない. 統語規則に特殊な活用の規則を加え, 「ない」のような特殊な形容詞に接続する付属語連鎖を解析するとき以外は, その規則を受容した付属語連鎖は検出する.

### 2.3 Bisonによる統語規則の実装

Bisonは, 統語規則をもとに特定のフォーマットで表記した構文規則と, 意味規則をもとにC言語で表記したアクションをソースコードとして, C言語で表記した構文解析関数に変換する. その関数と別に作成した字句解析関数を利用して付属語パーザを構成した.

字句解析関数は平仮名文字列をローマ字に変換して構文解析関数に送る. 構文解析関数はローマ字の文字列を構文解析する. ローマ字に変換するのは, Bisonの構文規則では平仮名文字を使用できないためである.

構文解析関数は付属語連鎖の統語規則を適用して構成する. 構文規則は分かりやすく表記するために, 平仮名で表しているが, それをソースコードとしてプリプロセッサでローマ字に変換したものをBisonでコンパイルするという方法をとっている.

2.2節にある統語規則の例(b)はBison内の構文規則では, `nai_type`, `ta_type`を各々「ない」助動詞, 「た」助動詞を表す非終端記号として

```
nai_type : なかつ ta_type ;
```

のように表記する.

動詞は活用のタイプによって活用が異なる. 例(a)ではBisonの構文規則は活用のタイプごとに必要である. たとえば, カ行五段活用の場合は,

```
k5 : 'k' z5 ;
```

```
z5 : あ nai_type ;
```

と表す. これは五段活用の動詞それぞれについて規則

を作らないようにするために, カ行五段活用の規則では, 他の五段活用と異なっているカ行の子音である'k'に五段活用で共通の活用形が続くようにした. ローマ字に変換したものを構文規則にするので, 子音を用いた表記が可能である.

アクションはその構文規則が成立したときに実行するC言語のステートメントで, 各々の構文規則に対して用意する. アクションでは,

- (1) 意味規則の処理を行う
- (2) 成立した規則中の付属語の数を数える
- (3) 仮名遣いの誤った付属語の構文規則が成立すれば, その付属語連鎖を検出する
- (4) 動詞に接続する付属語連鎖の規則であれば, 動詞の活用形を決定する
- (5) 名詞に接続する付属語連鎖の規則であれば, 名詞に接続する格助詞の種類を決定する
- (6) 付属語連鎖の語尾であれば, 文節の格を決定する
- (7) 成立した規則を構成する付属語のタイプを登録する

などの処理を行う.

Bisonが構文規則を変換して生成した構文解析関数は, LALR(1)パーザである. LALR(1)文法の構文規則は曖昧であってはいけない. Bisonは構文規則をBNF形式で表記するが, 曖昧な構文規則は許さない. 統語規則を構文規則にするにあたって, 曖昧な規則は1つを残してそれ以外を削除することで曖昧性を解消した. 規則の曖昧性が生じた付属語列の例を付録Cにまとめた.

曖昧な統語規則に受容される文字列がある場合, 適用する規則を1つに絞るということは, 文字列の解釈を文の内容を考えずに1つの解釈に絞り込むことである. 誤った単語分割を行った場合は付属語連鎖の意味属性の与え方を誤ることになり, 意味規則において検出過剰あるいは検出洩れなどの問題を引き起こす可能性がある. この問題に関しては後で言及する.

### 2.4 付属語連鎖における意味規則

付属語にはそれぞれ意味があり, 日本語のニュアンスを表現するために, 非常に大切な働きをしている. 日本語の統語規則では, 付属語どうしが接続することを比較的柔軟に許しているために, 付属語の意味属性が矛盾したり, 重なったりすることがある. また, 2.1節の例(vii)のように付属語によってはある意味属性が後に来る(この例では, 逆接)ことを期待させる呼応表現もある. つまり, 付属語連鎖の誤りを検出するには, 前節の統語規則に加えて, 矛盾, 冗長性, 呼応の

破れ、語用法の誤りを解析する意味規則が必要である。

そこで、意味属性の矛盾、冗長性、呼応の破れ、語用法の誤りを検出する意味規則を作成し、付属語パーズで解析することを試みた。この規則により検出された付属語連鎖は科学技術系の論文としては再考を要する表現であると見なす。

### (1) 意味属性

意味属性は助動詞、一部の助詞に与えた。意味属性の分類は、文献 11) を参考に次のようなものを用意した。

- テンスは次の 2 つに分類する。  
“意思・未来”，“過去”
- ボイスは次の 3 つに分類する。  
“受身”，“使役”，“可能”  
“可能”は受動態から発展したもの<sup>11)</sup>なのでこの中に入れた。
- アスペクトは次の 3 つに分類する。  
“現在進行”，“補助アスペクト”，“その他のアスペクト”  
アスペクトとは動作あるいは状態の完了、推移、開始を表す表現のことをいう。“現在進行”はテンスに分類されるという考え方もあるが、動作の推移状態を表すために、ここではアスペクトに入れた。
- ムードは次の 13 個に分類する。  
“断定”，“助動詞に接続する断定”，“準体助詞に接続する断定”，“限定”，“義務”，“強意”，“丁寧”，“推量”，“伝聞”，“推定”，“様態”，“その他のムード”，“命令”  
断定・推量・勧誘・命令を表す表現を総称してムードという。
- “みとめかた”は次の 4 つに分類する。  
“疑問”，“否定”，“逆接”，“欲求”  
ある事実の否定、疑問を表す表現を“みとめかた”という<sup>11)</sup>。“欲求”も事物をどう認めているかを述べているので“みとめかた”の中に入れた。
- “やりもらい”はこれ 1 つの分類とする。  
だれのために行われる動作であるのかを表すための特別な言い方を“やりもらい”という<sup>11)</sup>。
- “もくろみ”はこれ 1 つの分類とする。  
ある動作が何のために、何をもくろんで行われるのかを表す表現を“もくろみ”という<sup>11)</sup>。

断定の助動詞「だ」は接続する語によって意味属性を分けた。名詞・形容動詞に接続する断定の助動詞は“断定”の意味属性を、「みたいだ」の「だ」のように、助動詞の一部になっている断定の助動詞は“助動詞に

接続する断定”の意味属性を、「のだ」の「だ」のように、準体助詞に接続する断定の助動詞は“準体助詞に接続する断定”の意味属性を持つ。

“現在進行”は、現在の進行状況を表す意味属性で、「している」がある。“補助アスペクト”は、前に存在する動作あるいは動作の状況を表す語の意味を強調する意味属性で、「ところだ」がある。“その他のアスペクト”は動作の過去から現在への進行状況、現在から未来への進行状況、完了、開始などを表す付属語の意味属性である。「してくる」、「していく」、「してしまう」などがある。

“その他のムード”は断定的に述べる意味を持つ付属語の意味属性である。「わけだ」、「ものだ」、「はずだ」がある。

“やりもらい”の意味属性を持つものには「してあげる」、「してもらう」などがある。“もくろみ”の意味属性を持つものには「しておく」、「してみる」がある。

### (2) 意味規則の種類

意味規則は次の 4 つの規定で構成した。

#### (a) 後規定の誤り

たとえば、

- (正常に) 動作するはずであろう。

の例では、「はずだ」はものごとを断定的に述べるときに使用するから、推量の助動詞「であろう」に接続するのはおかしい。1 つの表現の中では“その他のムード”と“推量”は矛盾する。「はずだ」は、“推量”の意味属性をとまなわないように後規定しているという。

このように、1 つの付属語連鎖の中で、特定の意味属性を持つ付属語が、ある付属語の後にあるときに、後規定の誤りと判断する。

#### (b) 前規定の誤り

たとえば、

- (その民族は) 移住するのであるまでに、

の例では、語の用法上、強意の準体助詞「の」は付属語連鎖内の最後尾の準体助詞にならないと不自然になる。しかし、準体助詞「まで」は意味属性を持たないために、「の」の後規定で「まで」を禁止することはできない。意味属性を持たない付属語に対する意味規則として前規定を用意した。準体助詞「まで」は「まで」の前に強意の準体助詞「の」存在することを許さないように“強意”の意味属性を前規定しているという。

1 つの付属語連鎖の中で、特定の意味属性を持つ付属語が、ある付属語の前にあるときに意味属性の誤りと判断する。

#### (c) 二重禁止の誤り

たとえば、

- 開発してしまってきた。

の例では、「してしまう」という完了の意味の助動詞と「してきた」という過去から未来に動作が進行していることを表す助動詞が1つの表現に存在するのはおかしい。「してしまう」も「してくる」も動作の推移状態を表す意味属性の“その他のアスペクト”を持つ。1つの表現の中で2つの付属語が“その他のアスペクト”を表すと、表すものが同じならば冗長になり、違うものを表せば矛盾する。“その他のアスペクト”の意味属性は二重禁止であるという。

二重禁止はある意味属性を持つ付属語と同じ意味属性を持つ付属語が、1つの付属語連鎖の中に同時に存在するときに意味属性の誤りと判断する。

(d) 後必要規定の誤り

たとえば、

- 問題ではある。

の例では「ではある」は“逆接”の意味属性を持つ付属語と呼応がある。「ではある」は後に“逆接”の意味属性を持つ付属語がともなう必要があり、後必要規定しているという。

後必要規定は1つの付属語連鎖の中で、特定の意味属性を持つ付属語が、ある付属語の後にないときに意味属性の誤りと判断する。

### (3) 意味規則の例外

付属語には2つ以上の独立した意味属性を持つものがある。そのような付属語の意味属性には、科学技術系の論文でよく使用されていると思われる意味属性を採用した。

たとえば、“意思・未来”、“推量”の2つの意味を持つ助動詞「う」は、“意思・未来”の意味属性を持つものとする。“受身”、“可能”、“尊敬”の3つの意味を持つ助動詞「れる」は、“受身”の意味属性と持つものとする。

“過去”、“否定”の意味属性は、便宜上二重禁止を行わない。論文の中では強調のために以下のような表現を頻繁に使用するためである。

- 対処しなかったはずがない。(“否定”の二重禁止)
- 対処したからであった。(“過去”の二重禁止)

### (4) 自立語の意味属性

可能動詞は“可能”の意味属性を持つ。可能動詞を含めた状態動詞を表す動詞は動作を表さないのので、“受身”、“欲求”、“意思・未来”の意味属性を後規定する。たとえば、状態動詞「紅潮する」について考えると、助動詞「よう」を“意思・未来”として使用する場合は、

- 紅潮しよう。
- 紅潮したい。

という表現はおかしい。このように自立語の一部にも意味属性を与えた。

### 2.5 意味規則の実装

意味規則はアクションとして実装する。実際には、アクションから関数を呼び出して処理をしている。

意味規則を実装するにあたって、意味属性を登録する2種類の変数を用意した。後規定と二重禁止と後必要規定のために使用される意味属性変数、前規定のために使用される前規定変数である。前規定変数、意味属性変数に規定する意味属性を登録し、基本的には登録が2度目の場合に誤りと判断する。

付属語の規則に対応したアクションの中で、その付属語が規定する意味属性の前規定、後規定、二重禁止、後必要規定をする関数を呼ぶ。規定を行う関数では、以下の手続きをする。

- 二重禁止をする関数

(手続き a) 前規定変数に付属語の意味属性が登録されていれば、誤った付属語連鎖とする。

(手続き b) 意味属性変数に付属語の意味属性が登録されていれば、誤った付属語連鎖とする。

(手続き c) 意味属性変数、前規定変数に意味属性が登録されていない場合は、意味属性変数に意味属性を登録する。

意味属性を持つ付属語は、つねにその意味属性の二重禁止をしている。ただし、“否定”と“過去”の意味属性に関しては、(手続き b)を行わない。

- 前規定をする関数

(手続き d) 前規定変数に規定する意味属性を登録する。

- 後規定をする関数

(手続き e) 意味属性変数に規定する意味属性が登録されていれば、誤った付属語連鎖とする。

- 後必要規定をする関数

(手続き f) 意味属性変数に規定する意味属性が登録されていなければ、誤った付属語連鎖とする。

前規定、後規定などでは付属語の前後関係を知らなければいけない。本論文の統語規則におけるLALR(1)法のアルゴリズムでは、付属語連鎖内の最後尾の付属語から適用する。そのために、付属語パーザのアクションは付属語連鎖内で最後尾の付属語の規則から適用する。前規定、後規定、後必要規定は意味属性が変数にすでに登録されているのか、されていないのかをチェックすることで実現できる。

たとえば、2.4節の例(a)の「はずだ」が“推量”の意味属性を持つ「であろう」を後規定する場合を考える。まず、「であろう」を受容する。“推量”を二重禁

止する関数を呼ぶ。そして(手続き a),(手続き b)の処理を行う。そのとき前規定変数にも意味属性変数にも,“推量”は登録されていないので,(手続き c)を行う。“推量”が意味属性変数に登録される。次に,「はずだ」を受容する。“推量”を後規定する関数を呼び出す。(手続き d)を行う。“推量”が意味属性変数に登録されているので,意味属性の誤りとなる。

### 3. 付属語パーザの機能

#### 3.1 nspell の構成と付属語パーザ

nspell は,論文推敲に必要な情報を書き手に与えるシステムである。図 1 に nspell の機能図を示す。付属語パーザを利用した機能には,形態素解析,単語の検索,個別辞書作成がある。

形態素解析では,付属語パーザは自立語の辞書引きの結果得た自立語の品詞と自立語に続く平仮名文字列を引数として呼び出す。付属語パーザは,引数の品詞に接続する付属語連鎖として,引数の平仮名文字列は適切かどうかを判断する。付属語連鎖の解析結果は分割パターンの絞り込みや文節間の相関関係のチェックなどを使用する。

単語の検索は文字列の検索と違って,単語の品詞を考慮して検索する。普通,用言を検索する時には語幹を検索する。しかし,動詞などは語幹が平仮名で 1 文字か 2 文字であることが多く,それを検索すると意図しない単語まで検索してしまう。付属語パーザを利用して,品詞に適合した活用語尾とそれに続く付属語に接続するものだけを検索すれば,ほとんど目的とする単語だけを検索することができる。

辞書には,ユーザ辞書,システム辞書,個別辞書の 3 種類があり,ユーザ辞書はユーザがエディタで文章を編集中に登録・削除する辞書である。システム辞書には一般ユーザは登録・削除できない。個別辞書は,nspell が処理中にシステム辞書やユーザ辞書に登録されていない単語を自動的に登録してできる辞書である。nspell は,自立語の文字列とその品詞情報を書いたものを辞書とする。

個別辞書への自動登録は次のように行う。自立語が登録されていないと評価した仮文節文字列(漢字・片仮名文字列+平仮名文字列)中の漢字・片仮名文字列を仮に自立語とする。その自立語の活用情報を仮定して,その後続く平仮名文字列をその活用形の語に接続する付属語連鎖としてふさわしい場合には,単語の文字列として漢字・片仮名文字列を,その活用情報として仮定した活用情報を,個別辞書に登録する。この方法では,用言ならば語幹に,体言ならば単語中に,

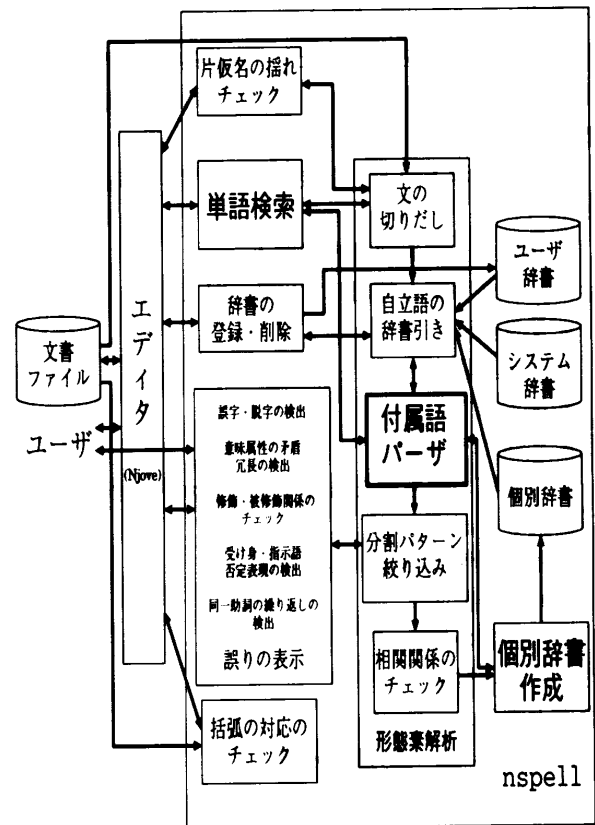


図 1 nspell の機能図

Fig. 1 Function structure of nspell.

平仮名の含まれた単語は登録できない。仮文節中に平仮名だけの文節が含まれる場合も,その文節中の自立語は登録できない。また,平仮名文字列によっては,品詞が形容動詞であるか名詞であるかの区別がつかないことがある。このような単語はあらかじめシステム辞書やユーザ辞書に登録しておく。

個別辞書は文書ごとに作成する。2 度目に同じ文書进行处理する場合は,個別辞書を用いて解析を行うと,自立語が辞書に登録されていないために,形態素解析が失敗することが少なくなり,付属語パーザの能力がより生かされる。

#### 3.2 nspell の機能

論文推敲支援システム nspell は文書作成エディタ Njove の一機能として動作している。Njove は福井大学工学部情報工学科において,Emacs とほぼ同等の機能を持つ JOVE を,完全に日本語に対応するように改良したエディタである。Njove は TeX による科学技術系の論文の作成支援機能や C 言語のプログラムの作成支援機能が強化されている。nspell は,Njove の科学技術系の論文の作成支援機能のひとつとして利用されている。エディタから子プロセスとして nspell

を起動し、その出力をもとに、エディタで編集集中に問題点へカーソルを飛ばしたり、nspell用の辞書への単語の登録・削除をエディタ内で行うなど、効率の良い機能を実現している。

nspellは以下のような項目について、問題点の検出を行う。

- 修飾・非修飾関係の問題点
- 同じ格助詞を使用している、連続した2文節
- 受身表現
- 否定表現
- 指示語
- 文体の不一致
- 句読点の打ち方を再考する必要がある表現
- 誤字・脱字
- 意味属性の誤りがある付属語
- 単語の検索
- 片仮名の揺れ
- 括弧の対応

などがある。

### 3.3 nspellの形態素解析

nspellの形態素解析は、局所的総当たり法<sup>12)</sup>を参考にしている。日本語は単語を構成する文字数が少なく、単語と単語が分けて表記されないのが曖昧性が生じやすい。総当たり型の形態素解析<sup>13)</sup>では、単語の分割パターンをすべてあげ、その分割パターンを絞り込むという操作をすることで、解析精度を向上させる。局所的総当たり法は総当たり型の解析であるが、平仮名とそれ以外の字種との境界で文節が区切れやすいことと自立語連鎖と付属語連鎖が必ず接続することを利用して、字種境界を利用しないものより、もっともらしい表現を残したまま単語分割パターンを減らし、ある程度高速な処理を実現している。

nspellの形態素解析には、局所的総当たり法とは異なった2つの特徴がある。1つは、誤った表現があることを前提に行うことである。もう1つは、付属語パーザを利用していることである。

従来の解析では誤った表現は、未登録語として処理をしていた。nspellの形態素解析では、解析に失敗した場合に、次の平仮名とそれ以外の字種との境界までを誤った文節であると見なす。平仮名とそれ以外の字種との境界が、文節の区切れ目になりやすいことを利用している。

### 3.4 付属語パーザの解析情報

付属語パーザは平仮名文字列を解析して、さまざまな情報を出力する。平仮名文字列が統語規則に受容された場合、付属語パーザは、文節の解析が終了した時

点で次のような情報を出力する。

- 自立語が用言の場合はその活用形
- 自立語が体言の場合は接続する助詞
- 文節の格に関する情報
- 仮名遣いに関する情報
- 構成する付属語の数
- 付属語の意味属性変数
- 付属語連鎖の正誤情報
- 付属語連鎖を構成する付属語のタイプ

付属語パーザの出力情報は、主に検出対象表現であるかの判断と、形態素解析の分割パターン絞り込みの評価の基準として使用する。分割パターン絞り込みでは、分割パターンの文節数や検出対象表現であるかどうかという情報だけでなく、動詞の連用中止形や、名詞に付属語が接続しない文節などが分割パターンに含まれるかを判定要素としている。

仮名遣いに関する情報とは、付属語の中で「づつ」や「けれど」などが使用されている場合、「づつ」ではなく「ずつ」であるとか、「けれど」ではなく「けれども」が正確な表記である、などの情報である。統語規則に「づつ」や「けれど」などの規則を加え、その規則が成立すると、仮名遣いの誤りとして登録する。

文節の格とは他の単語を修飾する関係を表す文節の性質<sup>11)</sup>をいうが、付属語パーザの出力する文節の格に関する情報とは、主格、目的格、連体修飾格、連用修飾格などの情報である。文節の相関関係チェックルーチンで、文節間の修飾・非修飾関係のチェックを行うときに使用する。

体言に接続する助詞の情報とは、体言に接続する助詞の種類を表す情報である。文節の相関関係チェックルーチンで、同じ助詞の接続している連続した2文節を検出するために用いる。

付属語の数は付属語連鎖が長すぎないかを調べるために用いる。科学技術系の論文は、簡潔な表現である必要がある。付属語連鎖の数は7個までとし、それ以上のものは長すぎる付属語連鎖として、検出する。

## 4. 推敲支援システムによる解析例と分析

付属語パーザを組み込んだシステムによって実際の文章の解析を行った。この目的は、付属語パーザの検出した表現が実際に推敲の必要な情報であるかどうかの評価を行うことである。nspellでコーパスを検索して、付属語の統語規則と意味規則が検出した文節を取り出した。さらに、人による手作業での誤りの検出と比較検討するために、同じ文章の問題点を手作業で点検・推敲した。

表1 付属語パーザによる処理の結果

Table 1 Result of estimation of agglutinated word parser.

群	コーパス 番号	統語規則		意味規則		手作業
		検出	合致	検出	合致	
A	(1)	1	0	0	0	0
	(2)	0	0	0	0	0
	(3)	0	0	0	0	1
	(4)	0	0	0	0	0
B	(5)	3	0	0	0	0
	(6)	1	1	0	0	2
C	(7)	1	1	0	0	3
	(8)	1	1	0	0	1
	(9)	1	1	0	0	1
D	(10)	0	0	5	2	7
	(11)	3	2	3	0	3
	(12)	2	1	5	0	3
E	(13)	2	1	0	0	1

文章中の引用文や括弧の内部は名詞と見なし、引用文と括弧の内部は処理対象外とした。

表1の一番上のコーパスのA群は、科学技術系の論文の例として工学系の論文、Bは翻訳の文章の例として工学系の論文の翻訳、Cは学生の論文の例として北陸支部の発表論文集に載っていた論文、Dは国語学の論文の例として国語学研究所の論文、Eは口語体の文章の例としてノーベル賞講演を取り上げた。

#### 4.1 システムの正解検出

統語規則の検出した表現でかつ手作業で検出した結果と一致した表現をあげる。検出した表現の推敲した結果を ⇨ の後に書く。前後の文脈を括弧の中に示す。

- 科学技術系の論文には不適切だと感じた表現。(2件)  
(メタ言語を)用いての(記述でなければ) ⇨ 用いた(詳細については)講演にて(述べる) ⇨ 講演で
- 古語的な表現である。(2件)  
(環境を通して)さざ波のごとく(広がるのである)  
⇨ さざ波のように  
(約束による)制度のごとき(ものという)  
⇨ 制度のような
- 誤字・脱字(3件)  
(最大値に)近づくことになる。  
⇨ 近づくことになる。  
(打電すると)いうとよ(特殊な事情)  
⇨ というよ(代表者を)選出してゆき、⇨ 選出していき、
- 口語的な省略(1件)  
(可能性があるかどうかを)調べてみてはと、  
⇨ 調べてみてはどうですか?」と

統語規則の検出では、誤字・脱字、科学技術系の論文としてふさわしくない付属語、口語的な省略表現の

検出が正確に行われている。

次に、意味規則の検出した表現で、手作業で検出した結果と一致した表現をあげる。

- 表現が厳密でないと感じた表現。(2件)  
(再編成された事実を)示していよう。  
⇨ 示していると考えられる。  
(省いたことによって)知られよう。

⇨ 知ることができる。

意味規則の検出では、国語学の論文に特有の「う」を“推量”の意味で用いた表現が検出されている。手作業で誤りとして検出したものと一致したこれらの表現は、文脈を考えると、表す意味はよく分かるが表現として厳密でないと思われる。

#### 4.2 システムの検出過誤

##### (1) 過剰検出(第2種の過誤)

システムが検出した表現のうち、手作業による検出では問題ないと判断した表現をまとめる。

- 統語規則によるもの。  
実証されるや、  
変更する、という  
行われるか、という  
記述、である。  
言えども  
量子力学のみならず、  
経済計画のみならず、
  - 意味規則によるもの。  
いえよう。(9件)  
できよう。(1件)  
必要があろう。(1件)  
発音されていようと、(1件)  
参照されたい。(1件)
- 第2種の過誤は、統語規則による検出表現は、現代語で使用される古語を用いた強調表現、読点の打ち方が特殊な表現が、意味規則による検出表現は、「う」を“推量”の意味で用いた表現が多く過剰検出されている。

##### (2) 検出洩れ(第1種の過誤)

手作業で誤りとして検出された表現の中でシステムでは検出されなかった表現を次に示す。前後の文脈を括弧の中に示す。

- 科学技術系の論文には不適切だと感じた表現。(3件)  
(言語能力が著しく)進歩してくる。  
⇨ 進歩し始める。  
(必要であることが)分かりつつある。  
⇨ 分かってきた。  
(記述されたという事実を)説明したものであろう



か? ⇔ 説明するべきであろうか?

- はっきりしない表現である。(6件)
  - (有力な手がかりと) なるはずのものである。
    - ⇔ なるべきものである。
    - (実在の言語が) 生ずるはずはない。⇔ 生じない。
    - (少し突っ込んで) 考えてみたい。⇔ 考えたい。
    - (おおよそのところは) 挙げえたつもりである。
      - ⇔ 挙げたつもりである。
    - (「文」の内容を) 規定することになる。
      - ⇔ 規定する。
    - (指定しなければならぬ) ものであるはずだ。
      - ⇔ ものだらう。
  - 漢字でなく平仮名である。(1件)
    - (位相特性は) 図2の様に(なる。) ⇔ 図2のように
  - 受身表現の使用方法がおかしい。(1件)
    - (検索手段として) 用いられている。⇔ 用いている。
    - 第1種の過誤には、文脈を考えなければ曖昧な表現であることが分からないような表現、助動詞「はずだ」を含んだ表現、助動詞「ように」を「様に」と漢字で書いた表現があった。

## 5. 考 察

出版されている科学技術系の論文および、翻訳の論文、国語学の論文を、本論文で提案した付属語パーザを組み込んだ推敲支援システム nspell で処理した結果、科学技術系の論文では検出された表現は少なかったのに対して、翻訳された論文および、国語学の論文は科学技術系の論文では、普通用いられない付属語を含んだ表現や口語的な省略表現が検出された。科学技術系の論文で検出された表現は、4つの論文でわずか1つであったが、国語学の論文は、3つの論文で18個もあった。これは付属語パーザが、科学技術系の論文にはあまり存在しない付属語連鎖だけを検出していることを示している。

口語表現の例としてあげたノーベル賞講演の文章には省略表現があった。文脈上理解できるが、省略部分が大きく、曖昧な表現である。これは、口語的な表現の特徴であると思われる。

国語学の論文には、科学技術系の論文では使用しないような表現が多かった。これらの表現を検出した原因は、意味属性の曖昧性解消のために、助動詞「う」、「れる」の意味属性を、科学技術系の論文でよく使用される意味属性に特定していることと、古語の付属語を統語規則に入れなかったことが考えられる。たとえば、「いえよう」、「できよう」の中にある、助動詞「う」の意味属性は“意思・未来”ではなく“推量”で、「参

照されたい」の中にある、助動詞「れる」の意味属性は“受身”ではなく“尊敬”である。「いえよう」、「できよう」などを問題点とするかどうかは、書く人の好みの問題だと思うが、ユーザに検出理由を明確に伝えるならば、検出対象としても問題はないと思う。ユーザに分かるようなメッセージを表示することが必要である。

手作業で誤りとして検出した国語学の論文中の表現自体は、日本語としては正しい。しかし、それらの表現は意味の厳密さより、文脈の流れから自然になることを重視して用いられていたために手作業では誤りとして検出した。科学技術系の論文では用いない方がよいと思われる。

手作業では3つの「はずだ」を含む表現が検出されている。助動詞「はずだ」を含む表現は科学技術系の論文では使いにくい表現である。今は検出対象になっていないが、意味が曖昧になりやすい付属語を含む表現も意味規則で検出対象にすべきだと考える。

曖昧な統語規則は1つを残してそれ以外を削除することで曖昧性を解消したが、今回の評価結果では過誤を生まなかった。曖昧性を持つ統語規則を片方に絞っても、文字列が受容するかどうかには問題がないことになる。しかし、統語規則の曖昧性を持った付属語間の意味属性が異なった場合、その付属語の意味属性をどちらのものにするのかは問題であり、実験結果では支障をきたさなかったが、種々の過誤を生む原因になる。

これからの主な課題として、(1) 統語規則や意味規則で検出された理由を、ユーザが理解できるように表示する、(2) 1つの付属語連鎖内にある意味上の区切り目を考慮する、(3) 2文節以上の相関を調べる、(4) 曖昧な統語規則に受容される付属語連鎖の2つ以上の分割方法を持ちながら次の解析に移れるようにする、などがある。

まず、理由を表示することに関しては、統語規則や意味規則で検出された理由が明確でない限り、書き手はその情報をどのように利用すればよいかを判断できない。分かりやすいインタフェースが必要である。

付属語連鎖の意味は、準体助詞「から」「か」および引用の格助詞「と」で区切れる。例として、「問題であるからだろう」がある。「問題であるから」と「だろう」は意味的には区切れている。「だろう」は“推量”の意味属性を持つ。“断定”の意味属性を持つ「である」が同じ表現の中にあると意味的に矛盾する。しかし、「から」で意味が区切れるので、問題のない付属語連鎖になっている。正しい検出を行うためには、付属

語連鎖内にある意味の区切り目を考慮する必要がある。

現段階では1文節の中だけの呼応関係しか調べていない。文脈や文の構造を解析した結果、初めて分かるような問題点は検出しない。つまり、格助詞の用法の誤りや「まず、～ない」のような呼応表現は検査対象になっていない。また、2.1節の例の(vii)に関しては、「問題ではある。しかし、～」のように使用することが多い。「ではある」は“逆接”の意味属性を後必要規定しているが、それは同じ付属語連鎖内である必要はないし、付属語である必要もない。上の例では、呼応する表現は「しかし」という自立語である。このように、付属語連鎖の誤った表現を検出するためには自立語の意味属性と2文節以上の相関を調べる機能を付加する必要がある。また、文節の相関関係を利用することで構文上の誤りをチェックすることが可能になる。

曖昧な統語規則に受容された付属語連鎖のために、付属語連鎖の解析結果を2つ以上持ちながら、次の処理に移ることができるようにする必要がある。本研究では支障をきたさなかったが、これから2文節以上の意味属性を解析するときには、さらに厳密な意味規則が必要になるため、これは必要であると思われる。

## 6. おわりに

本論文では、付属語連鎖の統語規則と意味規則について述べ、それをシステムで実現する方法についてまとめた。日本語の付属語連鎖の統語規則は複雑な接続が可能で、誤る可能性の高い部分である。付属語パーザの評価実験の結果、前章のすべての群の検出結果を総計すると、付属語パーザの検出した表現と手作業で誤りを抽出した表現が一致したのは、手作業で誤りとして検出した文節総数の45.5%であった。手作業で検出された表現には文脈から初めて不適当だと分かるものも多く、1文節内の統語解析、意味解析でこれだけの結果が出たのは、科学技術系の論文で使用する1つの付属語連鎖の中に、統語的、意味的な特徴があり、付属語パーザの統語規則、意味規則がその特徴をとらえているためであると考えられる。

科学技術系の論文における付属語連鎖の意味が曖昧であったり冗長であったりすると、表現したいことが伝わりにくくなる。特殊な言い回しや回りくどい表現は不要で、簡潔で正確な表現が必要とされる。翻訳の論文、論文を書き始めてまもない人の論文、国語学の論文などは、日本語としては誤ってはいないが、科学技術系の論文で使用する付属語連鎖としては不穏当と思われる表現を持っている。

正しい表現とは大多数の人間が使用する表現のこと

である。ひとりの意見で、正しいとか誤りだとか決めつけることはできないかもしれない。本論文の手作業で検出された問題表現の検出は、日本語を使う科学技術系の研究者のひとりとしての1つの基準であり、この付属語パーザが著者の基準をどれほど満たしているのかを検証した。このような評価結果は、人によっては、多少の差異はあるかもしれない。

最後に、共著者の川口湊教授は1995年9月26日、国際会議出席中に不慮の事故で亡くられました。ご冥福をお祈りいたします。

## 参考文献

- 1) 佐伯哲夫：陳述一文末の構成一，文化庁，chapter 5, pp.95-117 (1973).
- 2) 首藤公昭，榎原斗志子：日本語の文構造のわく組を与える表現—機能カテゴリと接続ルール—，福岡大学総合研究所報（自然科学編），pp.1-52 (1983).
- 3) 小野芳彦：入門システムプログラム，*SuperASCII*, Vol.2, No.#3-#5 (1991).
- 4) 丸山 宏，荻野紫穂：正規文法に基づく日本語形態素解析，情報処理学会論文誌，Vol.35, No.7, pp.1293-1299 (1994).
- 5) 奥村 薫，脇田早紀子，金子 宏：日本語校正支援システムにおける校正知識獲得—同音意義語について—，第48回情報処理学会全国大会論文集 (3)，pp.119-120 (1994).
- 6) 荒木哲朗，池原 悟，塚原信幸，小松康則：マルコフモデルによるかな文と英語文の誤り訂正，第51回情報処理学会全国大会論文集 (3)，pp.29-30 (1995).
- 7) 下園幸一，菅沼 明，牛島和夫：日本語文章推敲支援ツール「推敲」における助詞「が」の抽出について，情報処理学会論文誌，Vol.35, No.8, pp.1652-1660 (1994).
- 8) 高橋善文，牛島和男：計算機マニュアルの分かりやすさの定量的評価方法，情報処理学会論文誌，Vol.32, No.4, pp.460-461 (1991).
- 9) 鈴木恵美子，武田浩一：日本語文書校正支援システムの設計と評価，情報処理学会論文誌，Vol.30, No.11, pp.1402-1412 (1989).
- 10) 芳賀やすし：日本語文法教室，教育出版 (1982).
- 11) 鈴木 (監修)：概説・現代日本語文法，桜楓社 (1989).
- 12) 宮崎正弘，大山芳史：日本語音声出力のための言語処理方式，情報処理学会論文誌，Vol.27, No.11, pp.1053-1061 (1986).
- 13) 福島俊一：専用ハードウェアを用いた形態素解析器の開発，情報処理学会論文誌，No.3, pp.579-590 (1995).

付録 A. 付属語の分類

文献 10) の接続分類をもとに付属語を分類した。

準体助詞	か, かどうか, など, の, だけ, のみ, ばかり [完了], くらい, から [理由], ごと, ずつ, ころ, ばかり [限定]
後置詞	について, にしては, にしても
その他の後置詞	にとって, にしては, にあたって, にしても, によって, において, として, とはいえ, とともに
助動詞	せる, れる, らしい, みたいだ, そうだ [様態], そうだ [伝聞], つもりだ, はずだ, わけだ, もようだ, ところだ, た, しむ, ある, たい, ない, でしょう, ましょう, ます, う, だ, だろう, かもしれない, ねばならない, べきだ, ようだ, ません, まい, ものだ, なければならない, です, ん, であろう
接続助詞	ば [仮定], ならば, まま, て, のに, ており,
動詞接続の接続助詞	ながら [並行], ながら [逆接], つつ
接続助詞, 終助詞	くせに, ところで, ところが, けれども, どころか, が [逆接], し, に, から, ので,
格助詞 1	が, と, に, より, を
格助詞 2	で, へ, と
提示助詞 1	は, も, さえ, こそ, しか, すら
提示助詞 2	でも, だけ [限定], まで, のみ, ばかり
列挙助詞 1	と, か
列挙助詞 2	だの, やら, や
終助詞	か, な, よ, ね
て+補助助動詞	ておく, てみる, ている, てしまう, てある, てない, ていく, てくる, てあげる, てくれる,

て+補助形容詞

ていただく, てください, てもらう  
てほしい, てはいけない, てもよい

付録 B. 付属語連鎖の統語規則の例

統語規則 1266 個のうちからいくつかを選んで示す。非終端記号の中で“:”に続くものは意味属性を表す。意味属性を同時に 2 つ以上持つ付属語の意味属性は, “+” でつないで表記した。意味属性の中には意味規則で使用しない意味属性もまじっている。非終端記号内の“;”に続くものは接続分類を表し, “,” は要素の区切り目を表す。“意思・未来”はそれで 1 つの意味属性である。

- (動詞) → 動詞の終止形, (終止形接続)
- (動詞) → 動詞の連用形, (読点)
- (動詞) → 動詞の連用形, 「ながら」; 接続助詞: 並行, (読点)
- (動詞) → 動詞の連用形, 「ながら」; 接続助詞: 並行, (提示助詞 1)
- (動詞) → 動詞の連用形, 「ながら」; 接続助詞: 逆接, (読点)
- (動詞) → 動詞の連用形, 「ながら」; 接続助詞: 逆接, (提示助詞 1)
- (動詞) → 動詞の連用形, 「つつ」; 接続助詞: アスペクト, (読点)
- (動詞) → 動詞の未然形, (「せる」: 使役)
- (動詞) → 動詞の未然形, (「れる」: 受身・可能・尊敬)
- (動詞) → 動詞の未然形, (「なければならない」: 義務)
- (動詞) → 動詞の未然形, (「ない」: 否定)
- (動詞) → 動詞の未然形, (「ぬ」: 否定)
- (動詞) → 動詞の未然形, (「う」: 意思・推量)
- (動詞) → 動詞の連用形, (「ます」: 丁寧)
- (動詞) → 動詞の連用形, (「たい」: 欲求)
- (動詞) → 動詞の連用形, (「そうだ」: 様態)
- (動詞) → 動詞の連用形, (「まい」: 意思・未来+否定)
- (動詞) → 動詞の連用形, (「て」; 接続助詞)
- (動詞) → 動詞の連用形, (「た」: 過去)
- (動詞) → 動詞の終止形, (「でしょう」: 丁寧+推量)
- (動詞) → 動詞の終止形, (「だろう」: 推量)
- (動詞) → 動詞の終止形, (「かもしれない」: 推量)
- (動詞) → 動詞の連体形, (「そうだ」: 伝聞)
- (動詞) → 動詞の連体形, (「べき」: 義務)
- (動詞) → 動詞の終止形, (「らしい」: 推定)
- (動詞) → 動詞の連体形, (「の」; 準体助詞: 強意)
- (動詞) → 動詞の連体形, (「みたいだ」: 伝聞)

- (動詞) → 動詞の連体形, (「ようだ」: 様態)  
 (動詞) → 動詞の連体形, (「つもりだ」: 意思)  
 (動詞) → 動詞の仮定形, (「ば」; 接続助詞)  
 (動詞) → 動詞の命令形, (句点)  
 (終止形接続) → (動詞接続の接続助詞)  
 (終止形接続) → (動詞接続の助動詞)  
 (終止形接続) → 「より」; 格助詞, 「も」; 提示助詞, (読点)  
 (終止形接続) → 「より」; 格助詞, 「は」; 提示助詞, (読点)  
 (終止形接続) → 「より」; 格助詞, (読点)  
 (終止形接続) → 「しか」; 提示助詞, (読点)  
 (終止形接続) → 「に」; 接続助詞, (読点)  
 (終止形接続) → 「に」; 接続助詞, 「は」; 提示助詞, (読点)  
 (終止形接続) → 「どころか」; 接続助詞, (読点)  
 (動詞接続の助動詞) → (「だけ」; 準体助詞: 限定)  
 (動詞接続の助動詞) → (「から」; 準体助詞: 理由)  
 (動詞接続の助動詞) → (「わけだ」; その他のムード)  
 (動詞接続の助動詞) → (「まで」; 準体助詞)  
 (動詞接続の助動詞) → (「くらい」; 準体助詞: 限定)  
 (動詞接続の助動詞) → (「ところだ」; 現在進行)  
 (動詞接続の助動詞) → (「はずだ」; その他のムード)  
 (動詞接続の助動詞) → (「ものだ」; その他のムード)  
 (動詞接続の助動詞) → (「もようだ」; 伝聞)  
 (動詞接続の助動詞) → (「ばかり」; 準体助詞: 限定)  
 (動詞接続の助動詞) → (「など」; 準体助詞)  
 (動詞接続の助動詞) → (「か」; 準体助詞: 疑問)  
 (動詞接続の助動詞) → (「かどうか」; 準体助詞: 疑問)  
 (「だけ」; 準体助詞: 限定) → 「だけ」, (準体助詞に接続)  
 (「だけ」; 準体助詞: 限定) → 「だけ」, (格助詞 1)  
 (「だけ」; 準体助詞: 限定) → 「だけ」, (「か」; 準体助詞: 疑問)  
 (「だけ」; 準体助詞: 限定) → 「だけ」, (「かどうか」; 準体助詞: 疑問)  
 (読点) → 読点  
 (読点) → ε

### 付録 C. 曖昧性を持つ付属語列

曖昧な規則を構成する付属語列を以下に示す。改行までが1エンタリである。エンタリの1行目は曖昧性を持つ規則が受容する文字列, 2行目, 3行目, 4行目は規則を構成する付属語列である。非終端記号内の‘, ’は要素の区切り目を表し, ‘;’に続くものは接続

分類を表す。接続分類だけが書かれている場合はその分類に入るすべての単語が適用される。数が多いために、この付属語列の前後に何が接続するかは示していない。

( ) は注意事項を書いた。3行目以降の付属語列を統語規則から削除した。

で

「だ」: 断定

「で」; 格助詞 2

たところ

「た」: 過去, 「ところ」: 補助アスペクト, 「だ」: 断定

「た」: 過去, 「ところで」; 接続助詞

でも

「でも」; 提示助詞 2

「で」; 格助詞 2, 「も」; 提示助詞 1

か

名詞, 「か」; 準体助詞

名詞, 「か」; 列挙助詞

と

名詞, 「と」; 列挙助詞

名詞, 「と」; 格助詞

ながら

「ながら」: 並行

「ながら」: 逆接

のに

「の」; 準体助詞, 「に」; 格助詞 1

「のに」; 接続助詞

ていただけ

「ている」: 現在進行, 「た」: 過去, 「だけ」: 限定

「ていただく」: やりもらい

は, も, さえ, しか, すら

提示助詞 1

「が」; 格助詞, 提示助詞 1

「を」; 格助詞, 提示助詞 1

(提示助詞は, 「が」「を」に接続する時には表記のう  
えでは, 提示助詞だけになる)

ないでも

「ない」: 否定, 「で」; 接続助詞, 「も」; 提示助詞 1

「ない」: 否定, 「でも」; 提示助詞 2

ばかり

「ばかり」; 準体助詞: 完了

「ばかり」; 準体助詞: 限定

まで

「まで」; 準体助詞

「まで」; 提示助詞 2

ので

「の」;準体助詞,「だ」:断定  
「ので」;接続助詞

(平成7年12月13日受付)  
(平成8年9月12日採録)



蓮井 洋志 (学生会員)

1970年生。1993年福井大学工学部情報工学科卒業。1995年同大学工学研究科情報工学専攻修士課程修了。現在、同研究科システム設計工学専攻。自然言語処理に興味を持つ。

電子情報通信学会学生会員。



川口 湊 (正会員)

1959年東京大学理学部物理学科卒業。1965年同大学院数物系博士課程単位取得退学。1965年東京大学助手、1976年福井大学教授。その間、1966~1971年米国ブルックヘブン

国立研究所客員研究員を務めた。理学博士。1995年ブラジルで国際会議に参加中、不慮の事故で逝去した。高エネルギー研究、高エネルギー分野での高性能超並列コンピュータなどの研究とともに、文書清書システム、論文作成支援システム、日本語論文推敲支援システムなどに強い関心を持ち、研究を進めてきた。日本語物理学会、情報処理学会、各会員であった。(共著者記す)



小倉 久和 (正会員)

1969年京都大学理学部物理学科卒業、1977年同大学院理学研究科博士課程修了。1979年高知医学大学助手、1986年同大学助教授。1988年福井大学助教授を経て、現在、同大工学部情報工学科教授。理学博士。学習、言語理解、ファジイ推論、遺伝的アルゴリズム、医用画像理解、医療情報処理など、知識情報処理を中心に研究を行っている。著書「図解総合医療情報システムの実際」(メジカルビュー社、共著)、「情報の論理数学入門—ブール代数から述語論理まで—」(近代科学社、共著)、「形式言語と有限オートマトン入門—例題を中心とした情報の離散数学—」(コロナ社)など。情報処理学会、電子情報通信学会、人工知能学会、日本医療情報学会、日本ファジイ学会、IEEEなど各会員。