

長い日本語文における並列構造の推定

黒橋 禎夫 長尾 眞

京都大学工学部 電気工学第二教室

要 旨

長い日本語文では複数の内容が並列的に述べられていることが多いが、それらの並列構造の正しい認識が困難であるために長い文の構文解析はほとんど失敗する。多くの場合、並列している部分は何らかの意味において類似している。そこで、まず文内の全ての文節対について類似度を計算し、次に並列の存在を示す表現の前後で類似度の総和が最も大きい2つの文節列をダイナミックプログラミングの手法によって求め、これを並列構造の範囲とする方法を考案した。並列構造としては、名詞句並列の他に連用中止法による述語句並列等を対象とした。180文に対し実験を行なったところ、この方法によってかなりうまく並列構造を推定できることがわかった。

A Method for Analyzing Conjunctive Structures in Japanese

Sadao Kurohashi Makoto Nagao

Department of Electrical Engineering, Kyoto University

Abstract

Parsing a long sentence is very difficult, since long sentences often have conjunctions which result in ambiguities. As the conjunctive parts in a sentence often appear in a similar structure, finding two similar series of words is essential in analyzing conjunctive structures. Similarities of all pairs of words are calculated. Then the two series of words which have the greatest sum of similarities are found by dynamic programming method. We deal with not only conjunctive noun phrases, but conjunctive predicative phrases created by "Renyoh chuushi-ho". We will illustrate the effectiveness of this method by the analysis of 180 Japanese sentences.

1 はじめに

機械翻訳を代表とする日本語情報処理は最近かなりの進展をみせている。しかし困難な問題はいくつも残されており、それらは未解決のまま放置されている。その1つに長い文の解析の問題がある。日英機械翻訳システムは数多く出され実用されはじめているが、漢字かなまじり文で50文字以上の文の解析は非常に困難であり、80文字以上の文の解析はほとんどが失敗するといわれている。

そこでその失敗の原因を考えてみると、「長い文の場合、1つの文節が係ってゆく先がいくつもありうるから」ということは誰にも分かる。しかし、それがなぜ解析の失敗につながるかということが明らかでない。我々は、その誤り原因の疑いを、採用している文法形式に向けてみる。今日に到るまでの過去30年間、言語の文法といえ基本的にはチョムスキーの提案した句構造文法にたよってきた。もちろんその改良として、属性をもたせた拡張句構造文法がいろいろと提案され、ユニフィケーションという操作を導入することによって種々の試みが行なわれて来ている。また一方では格文法という考えが導入され、語順にあまり影響されない解析が行なえるようになっていく。解析結果も句構造表現、格構造表現、依存構造表現、その他種々の形式が存在する。しかしこれらいずれの場合にも基本的な考え方は隣り合う複数個(ほとんどの場合、2個)の要素の関係をしらべるというものである。格文法の場合は少しはなれた要素を取り上げることもあるが、その場合でも特別なマーカー(日本語の場合、格助詞)のついたものを取り上げ、他のもう1つの要素(動詞)との関係を考えるというものである。

しかし言語表現はそんなに単純なものではない。いくつもの、しかもかなり離れたところにある要素同士が呼応しあうということはいくつもある。構文解析によって多くの可能な解析結果を出した(出せる)といった報告は、ある意味では文法が不完全だから排除すべき構造まで出してしまうわけで、決してほめたものではない。そのように多くの解析結果が出る主な原因は、隣り合う2要素の関係しか見ないところにあることを知るべきである。

さて、それではどのようなことを考えるべきか。1つは各要素に付属させる(意味)属性値をできるだけ精密なものにし、適切な規則だけが適用されるようにすることである。他は文中に広くひそかに存在する多くの要素同士の関係を同時に検出することである。文が長くなる主な原因は、1文中に多くの内容を並列的に述べよう

表 1: 並列のキーを示す語

(a) 名詞並列
[読点]・[中点]ともやかとかつだけで(は)なく およびまたはならびにあるいはもしくは
(b) 述語並列
の+に対し(て)とかかしが+にだけで(は)なく けれど(も)およびまたはならびにあるいはもしくは
(c) 部分並列
およびまたはならびにあるいはもしくは

注) '+'は連続する語であることを示す。

とするとところにある。並列的なものとして、ここでは並列名詞句、並列形容詞句などのほかに並列する文、即ち連用中止法による文の継続的接続も含めて考えている。そして多くの文において、これらの並列する部分は何らかの意味において類似しているのである。特に科学技術文章などにおける長い文ではそういったことが多い。したがって、このような並列する構造を正しく認識できれば、長い文も極端に短い形にすることができ、文の解析が正しくできる可能性が高くなる。

しかし、この類似性を従来方式の構文解析によって発見するというのでは問題の解決にはならない。また、人間はそれほど構文解析的な立場で考えながら並列性を含んだ長い文を書いているわけでもないだろう。したがって長い文の中に存在する類似した2つの単語列(日本語の場合、文節列)を上記に説明したような従来の文法規則の適用でない方法で発見することが出来る必要がある。本稿では、これを音声認識などで広く使われているダイナミックプログラミングによるマッチング法(DP マッチング)の考え方によって実現した。

2 日本語文における並列構造

まず日本語文における並列構造にどのようなものがあるかを整理しておく⁽¹⁾⁽²⁾。

1つには、表1(a)に挙げた語によって接続される名詞の並列がある。

- (i) ... 解析と 生成を ...
- (ii) ... 原言語の 解析と 相手言語の生成を ...
- (iii) ... 原言語を解析する 処理と 相手言語を生成する 処理を ...

これには上の例文のように、単独名詞の並列、修飾語を伴う名詞の並列、連体修飾節を伴う名詞の並列がある。ここでは、これらを全て含めて名詞並列とよぶことにする。

もう1つは、1文内に複数の述語¹があり、それらが並列的に、すなわちどちらが主とも従ともいいがたい平等の関係でつながっているという構造である。

(i) ...原言語を解析し 相手言語を生成する ...

(ii) ...解析では利用するが、生成では利用しない ...

この種の並列の存在は (i) のように連用中止法によって示される場合と、(ii) のように表 1(b) に挙げた接続助詞などによって示される場合がある。このような並列構造を述語並列とよぶことにする。

その他の並列構造として、述語並列から述語を除いたある一部分が並列につながっているというものがある。

(i) ...前者を解析に、後者を生成に ...

(ii) ...解析に、または生成に ...

これには (i) のように助詞が呼応している場合と (ii) のように明示的に並列構造を示す語 (表 1(c)) を伴う場合がある。このような並列構造を部分並列とよぶことにする。

また、各並列構造の存在を示す表現 (上の各例文の下線部分) を並列のキーとよぶことにする。

3 並列構造の推定の方法

本稿では並列構造の推定をつぎのような問題として扱う。

並列のキー²に対して、その前後のどの範囲が並列構造をなす2つの文節列であるかを推定する。

例えば、「...A の B と C の D を ...」という表現において、「と」によって'B'と'C'の並列が示されているのか、'A の B'と'C の D'なのか³、あるいは前後のさらに広い範囲までを含むのかを推定することになる。

この方法の概要を、図1の三角行列によって説明する。

前処理 まず入力日本語文を形態素解析し、自立語とそれに続く付属語を1つのブロックにまとめる (以後、この各ブロックを文節とよぶ)。図の各対角要素が1つの文節である。

文節間の類似度の計算 全ての文節の対について類似度を計算する。類似度としては自立語が同じ品詞である場合、同じ付属語を含む場合などにポイントを与える。

三角行列の (i,j) 要素の数字は i 番目と j 番目の文節の類

¹ 述語になり得るのは、動詞、形容詞、形容動詞、および、「だ」/「である」/「です」のいずれかを伴う名詞である。

² 表1の語や連用中止法などを手掛りとして機械的に取り出す。ただし、「・(中点)」による名詞並列は、ほとんどの場合その前後の単独名詞同士の並列を示していると解釈できるので、並列のキーとはしない。また、読点を伴わない連用中止の述語も、ほとんどの場合次の述語に結びつくと考えられるので並列のキーとはしない。

³ 'B'と'CのD'、'AのB'と'C'なども考えられる。

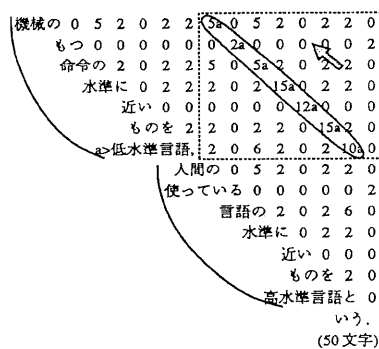


図 1: 並列構造の推定の例

似度のポイントを示す。

並列構造の範囲の推定 並列のキー (図では 'a>' のついた文節) の前後で、類似度の総和が最も大きい文節列の組を求める。これは、図の点線の範囲内において、一番下の行の1つの要素から出発して点線の範囲内の一番左の要素までの左上方向への要素の並び (以後、これをパスとよぶ) の中でポイントの和が最も大きいパスを求めることに対応する。この処理は DP マッチングの手法で行なう。そのようなパスが求められれば、そのパスの左側の対応する文節列と下側の対応する文節列が並列であると推定する。図の例文では、添字 'a' のついたパスが最高得点 64 を得るパスで、このパスに対応する2つの文節列、「機械の〜低水準言語」と「人間の〜高水準言語と」が並列であると推定される。

この処理について以下で詳しく説明する。

3.1 前処理

まず形態素解析⁵によって入力日本語文を単語単位に分割し、各単語の品詞、さらに活用する語の場合には活用形と原形を決定する。この解析は複数の解析結果を出す。並列構造の推定に関心があるので複合語などはなるべく一単語として扱う方がよいという理由から、以下の処理はこのうち単語数最少、自立語数最少の解に対して行なう。

次に分割された単語列を文節 (自立語とそれに続く付属語) にまとめる。ここでは、読点も付属語とみなし、読点の次に「または」、「あるいは」などの接続詞が続く場合にはそれらも同一文節の付属語とみなす。また、意味

⁴ 1文中に複数の並列のキーが存在する場合もあるので、これらをアルファベットを添えて区別している。

⁵ 益岡・田窪文法⁽³⁾を拡張したものを標準文法とする形態素解析プログラムを使用する。この際、機能的に助詞とみなせる「という」、「に対して」、「だけでなく」などは助詞として扱う。

的まとまりという観点から「サ変名詞＋する」、「サ変名詞＋を＋行なう」などは一つの文節にまとめ、自立語の品詞は動詞であるとし、「する」、「行なう」などは付属語とみなす。

3.2 文節間の類似度の計算

文節間の類似度は次のような5つの基準によって計算する。なお、活用する語については原形で比較する。

1. 自立語の品詞が一致している場合にポイント2を与える。以下のポイントは自立語の品詞が一致しているものについてのみ加算していく。
2. 自立語が同じ単語である場合、ポイント10を加える。この場合は、次の自立語の部分一致によるポイント、分類語彙表によるポイントは与えない。
3. 自立語が名詞である場合に限り、自立語の文字列が部分的に一致する場合、一致した文字列×2ポイント(最大10ポイントで打ち切る)を加える。
4. 各々の自立語について国立国語研究所の分類語彙表⁽⁴⁾のコードを調べ、コードの上位桁から下位桁にむかって連続する一致が3桁以上である場合、(桁の一致－2)×2ポイント(3の文字列部分一致のポイントと合わせて10ポイントで打ち切る)を加える。
5. 同一の付属語がある場合、付属語の一致一組につき3ポイントを加える。

例えば、「低水準言語＋」と「高水準言語＋と」では、2(品詞の一致)＋8(文字列部分一致: 4文字)で10ポイント、また「訂正＋し＋」と「検出＋する」では2(品詞の一致)＋2(分類語彙表コード: 3桁)＋3(付属語一致)で7ポイントとなる。

ここまでのポイントは、自立語の品詞の一致を前提としたものであるが、そうでない場合でも次のものには特別にポイントを与えた。

- 「サ変名詞＋する」などの文節は、ひとまとまりの動詞として扱っているが、その中のサ変名詞だけが「…解析、生成する…」、「…解析と生成を行なう…」のように並列構造をなしている場合がある。このような並列構造の範囲を正しく推定するためには、並列している文節(下線部分)の間に類似性を認めておく必要がある。そこで、サ変名詞を動詞として扱っている文節と、名詞として扱っている文節の間に2ポイントを与え、さらにサ変名詞同士について上の2～4のポイントを与える。
- 述語並列のキーとなっている文節は、それより後ろのいずれかの文節と対応しているが、それは必ずしも自

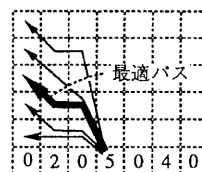


図2: 起点からの最適パス

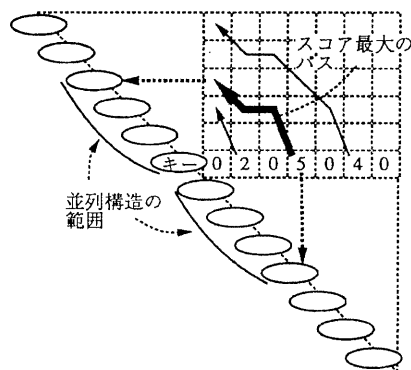


図3: 並列構造の範囲を示すパス

立語の品詞が同じ文節であるとは限らない(例えば「…が強力で、…ができる装置を…」)。そこで、述語並列のキーの文節と、それより後ろの述語となり得る文節との間に類似性を認めておく必要があるため、そのような文節間にポイント2を与える。

3.3 並列構造の範囲の推定

並列のキーの前後で、ある種の制限のもとで類似度の総和が最も大きい文節列の組を求め、それらを並列構造の範囲とする。このような文節列の組を、文節間の類似度のポイントを要素とする三角行列において以下のような方法で求める。

並列のキーの右上四角形部分を部分行列とよぶことにする(図1では点線内の行列に対応)。この部分行列の一番下の行の0でない要素を起点とし、そこから一番左の列の要素までのパスを考え、このうち後で説明する方法で計算されるスコアが最大のパス(以後これを最適パスとよぶ)を求める(図2)。この計算を部分行列の一番下の行の0以外の全ての要素を起点として行ない、各起点からの最適パスのうちで最もスコアの高いパスが並列構造の範囲を示すものであるとする(図3)。

日本語文の各文節は意味的に右側に係っていくので、最も重要なことは並列のキーの文節が後ろのどの文節に対応しているかということである。これは、本手法では

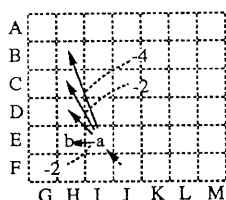


図 4: スコアの計算法

パスの起点をどの要素にするかという問題に対応する。しかしここでは、「並列のキーの文節と最も類似した文節はどれであるか」というような単純な 1 対 1 の対応を考えるのではなく、各々の文節の前の部分がどのように対応しているかという広い範囲の情報を総合的に調べる。これがスコアの高いパスを求めることに対応するが、このことによって妥当な並列構造の推定を行なうことができる。

以下にスコアの計算方法を具体的に説明する。

基本的スコア計算法 パスは起点となる要素から始まって各列の要素を 1 つずつ順につないだものである。この中で各要素をつなぐ部分を枝、枝の起点側の要素を始点、反対側の要素を終点とよぶことにする。枝としては、左横方向へのものと左上方向へのものだけを許す。あるパスのスコアは、基本的にはパス内の要素のポイントの和である。ただし、次のことを考慮する。

- 枝が左横方向である場合はその枝の終点のポイントはスコアに加えない。これは、「類似度を考慮する文節の対応としては 1 文節対 1 文節の対応だけを考える」ということを意味する。例えば、図 4 のパス「 $a \rightarrow b$ 」では、文節列の対応としては「E」と「H I」の対応を考えているが、文節の対応としては「a」による「E」と「I」の類似度のポイントだけをスコアに加え、「H」については対応するものがないと考える。

さらに次の減点を行なう。

- 枝が左横方向である場合は 2 だけスコアを減らす。枝が左上方向で始点と終点の行の差が 2 以上である場合は、(行の差 - 1) × 2 だけスコアを減らす (図 4)。これらは、「文節数が同じぐらいの、バランスの取れた並列構造がより自然である」ことを表現している。

ペナルティ 2 つの並列文節列の各々はその範囲内で 1 つの構造にまとめられることが多い。したがって、「... 装置は、生産と検査の自動化を ...」というような文で「～は、」という文節が「と」による並列構造

表 2: 区切りレベル

レベル	表現
5	述語並列のキーの文節「～は、」
4	「～(名詞並列を示すもの以外の助詞),」 「(副詞),」
3	「(活用語の連用形)」 「～は」
2	「～(名詞並列を示す付属語),」
1	「～,」 「～(名詞並列を示す付属語)」

表 3: ボーナスを与える表現

	名詞並列	述語並列
並列構造の最後の文節	など 等	ために ための という といった ようだ など 等
並列構造に続く文節	各～～種類 ～つ 組 対 両方 など 等	ことものとき方式 方法手法 など 等

の中に含まれるとは考えにくい。即ち「装置は、生産」と「検査の自動化を」が並列であるとは考えられない。このような現象を考慮するために、文節列の間にある種の区切りを示す要素が存在するとスコアを減じ、それにより、さらに広い範囲にまで並列文節列がのびてゆく可能性を小さくするようにする。そのために種々の区切り要素に対して表 2 のように 5 段階の区切りレベルを設定する。そして、並列のキーの文節の区切りレベルを基準とし、それと等しいか、あるいはそれよりも強い区切りレベルの文節を並列構造に含もうとする場合にペナルティを与える。ペナルティは、(区切りレベルの差 + 1) × 7 とする。ただし、区切りレベルが高い場合でも、同じタイプで、しかも並列のキーとはタイプの違う 2 つの文節が、対応して並列構造の中に現れる場合、すなわち、その 2 つの文節の対応を示す要素がパスに含まれる場合にはペナルティを免除する。ここで、文節のタイプが同じであるとは、2 つの文節において自立語の語彙以外が全て一致すること、すなわち、自立語の品詞と活用形、全ての付属語が一致することを指す。このような免除を行なうのは、「～は、～を行ない、～は、～を行なう。」のような文において対応関係にある「～は、」が並列構造の範囲を限定しているとは考えられないからである。

ボーナス 名詞並列の後の「など」のように、並列構造の直後にあって並列構造の範囲の推定に有用な表現がある。そこで表 3 のような表現が並列構造の最後の文節、すなわち起点の要素の下側の文節や、その後ろの文節にある場合、その起点からのパスのスコアにボーナス 6

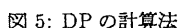


表 4: 解析結果の評価

文字数	情報科学辞典			JICST 抄録文			サイエンス			計
	30-50	50-80	80-149	30-50	50-80	80-144	30-50	50-80	80-139	
名詞並列のキーを含む文	8	13	12	8	11	13	5	5	11	86
正しい範囲を推定したもの	5	11	10	7	6	5	5	5	10	64(74%)
誤った範囲を推定したもの	3	1	2	1	5	8	0	0	1	21
並列のキーとして適当でないもの	0	1	0	0	0	0	0	0	0	1
述語並列のキーを含む文	6	16	18	4	15	14	3	7	11	94
正しい範囲を推定したもの	6	15	15	3	10	9	1	5	6	70(74%)
誤った範囲を推定したもの	0	1	2	1	5	5	2	2	5	23
並列のキーとして適当でないもの	0	0	1	0	0	0	0	0	0	1

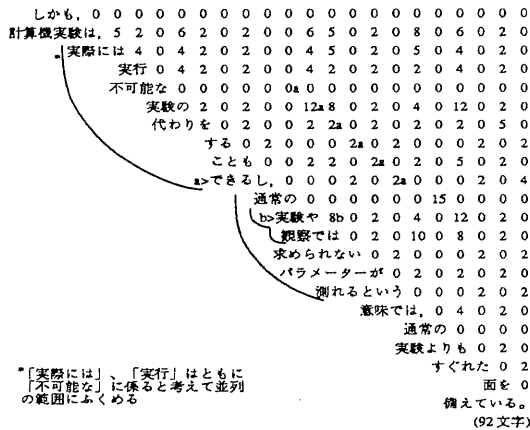


図 8: 並列構造の推定の例 (3)

スが有効に働いている。

4.2 定量的評価

180 文の解析結果について人手で評価を行なった。表 4 はその結果を名詞並列、述語並列それぞれについて文単位にまとめたものである。1 文中に同じタイプの並列構造が複数ある場合には、そのすべてに対して正しい構造を推定している場合を正解とした。

並列のキーの数としては、名詞並列のキーは 147 個でそのうち正しい並列構造を推定できたものは 121 個 (正解率 82%)、述語並列のキーは 119 個でそのうち正しい並列構造を推定できたものは 94 個 (正解率 79%) であった。

なお、並列構造としてはもう 1 種類、部分並列を考えていたが、実験に使用した 180 文中にはこのタイプの並列のキーは存在しなかった。

表 4 をみると JICST の抄録文の解析結果があまり良くない。これは、抄録文では限られた文字数で多くのこと

を記述しようとするために、人間が読んでも難解であるような文が多いためである。

4.3 失敗例と解決法

実験で正しい並列構造が推定できなかった文の具体例を示し (表 5)、解決の見通しのあるものについてはその方法を述べる。表 5 では、下線部分が並列のキーの文節および誤って推定した範囲においてそれと並列する文節、「」が誤って推定した範囲、『』が正しい範囲である。

1. 本手法では類似した文節間に適切なポイントを与えることがまず重要である。そこで、品詞、特に名詞を数詞、固有名詞、サ変名詞、普通名詞などに細分類し、類似度の与え方に差をつけることが考えられる。例文 1 の誤りは、サ変名詞「拡張」と普通名詞「困難」の類似度よりも「拡張」と「保守」のサ変名詞同士の類似度を大きくするというのを、他の要因とのバランスを保った上でうまく行なえれば解決できる。
2. 現在、意味的類似性は分類語彙表のみを用いて与えているが、これに加えて専門用語のシソーラスなどが利用できれば 1 と同じ理由から並列構造推定の精度が向上すると考えられる。例文 2 では「アクティブ・チャート解析法」と「HP SG」の類似度により大きなポイントが与えられれば正しい構造が推定できる。
3. 推定された並列構造が統語的にみて誤っている場合がある。例文 3 の推定された並列構造では「文法を」の係り先がない。この例では、「文法を」と「解析と」の間に動詞がないことから、並列構造の前の範囲は「自然言語の解析と」か「解析と」しかないことが分かる (もちろん、「～を～と、～を～と」のような助詞の呼応がないことを調べる必要がある)。このような処理は、推定された並列構造の情報を用いて構文構造を決定していく次のレベルの処理として実現することを

表 5: 解析の失敗例

例文 1: これら解析手法の共通した問題として文法規則が大きくなった場合の「規則の『拡張や 保守の』困難が」上げられる。(49 文字)
例文 2: ATR 自動翻訳電話研究所で開発された音声言語日英翻訳実験システム S L - T R A N S の日本語対話文解析部は、『「解析過程の制御が自由な アクティブ・チャート解析法と 単一化に基づいた語い・統辞的な文法的枠組みである」HP SG を』採用している。(113 文字)
例文 3: 「単一の文法を自然言語の『解析と 生成に』用いる双方向文法の 研究は、」計算言語学の上からも、機械翻訳や自然言語インタフェースといった応用面からも重要である。(73 文字)
例文 4: 実際、筆者たちは『「これを使って、重力相互作用が支配する」天体の運動について、高精度で高速の数値計算ができるデジタル・オレリーという専用コンピュータを製作している。』(81 文字)
例文 5: 一般に、生成アルゴリズムが完全であることは証明できるが、『「非文に対する 停止性や 出力する文の あいまいさの」上限について』保証がない。(62 文字)
例文 6: 『述語相当の慣用的表現である「 <u>述語慣用語と 機能動詞表現の</u> 」』二つに関して、それぞれの解析手法について提案する。(51 文字)

考えている。

4. 並列のキーが文のはじめであればあるほど、その後ろに並列のキーの文節と対応する可能性のある文節がたくさんあるので、並列構造の推定は難しくなる。例えば例文 4 のように文のはじめの連用中止が文末の述語と対応するような場合の解析は、バランスのとれた並列構造でないために非常に困難である。このような文を正しく解析しようとするれば「使って」と「製作する」の間の因果関係のような情報が必要であろう。
5. 微妙な表現であって、人間が読んでも曖昧であったり、専門的知識がなければわからないといったものがある(例文 5)。
6. 解析の失敗というよりも、本手法では本質的にうまく扱えない問題として例文 6 のような場合がある。このように、並列構造の前半部分に含まれ後半部分には対応するものがないような文節列がある場合、正しい並列構造を推定することができない。

例文 5 や例文 6 を正しく解析しようと思えば意味情報を使う以外に方法はないが、それは言語情報処理の最終

目標の 1 つであろう。

なお、実験した例文については失敗例がなかったが、「～、さらには」「～、例えば」のように挿入句的な形で存在する並列構造は、本手法ではうまく扱えない場合がある。この取り扱いについては別に考慮中である。

5 おわりに

以上に示したように、長い文の内部構造として多く存在する並列文節列が、文節列同士の類似性の発見という考え方でかなりうまく検出できることが分かった。この方法の導入によって多くの長い日本語文が正しく解析されるようになるだろうが、それでも解析に失敗する場合はまだまだ残る。しかしそのような失敗についても、深い意味解析に行かずに表層的な手係りによって解決できる場合はいろいろあると考えている。単純な文法規則しか導入せず、それで多数の解析結果が出たり、解析に失敗したらすぐ意味処理によって解決しようとするのは、いささか安易な考え方であると思う。意味素性を導入し、意味的整合性をチェックするという方法もよほど精密なものを作らない限り実際にはそれほど効果を発揮しない。他に有効な意味処理の方法は現在のところないのであるから、できるだけ構文的な現象を綿密にしらべることが必要なのである。これをひと口でいえば、1 つの文、あるいは複数の文の並びにおいて、できるだけ広い範囲の言語表現を同時的に調べるということであろう。

句構造文法、格文法、ユニフィケーション文法だけが文法なのではない。文中の多数の言語要素間にどのような関係性を認めるかということが文法なのである。その関係性の検出を具体的にどう実現するかは工学的手法の問題である。そしてこれには種々の方法をとることができるだろう。問題は今までに明確には認識されていなかった、文中における微妙な文法性の明確化であろう。

参考文献

- (1) 長尾, 辻井, 田中, 石川: 科学技術論文における並列句とその解析, 情報処理学会自然言語処理研究会報告, 36-4, (1983).
- (2) 首藤, 吉村, 津田: 日本語技術文における並列構造, 情報処理学会論文誌, Vol.27, No.2, pp. 183-190, (1986).
- (3) 益岡, 田窪: 基礎日本語文法, くろしお出版 (1989).
- (4) 国立国語研究所: 分類語彙表, 秀英出版 (1964).