

# 科学技術論文における並列句とその解析

長尾 真・辻井 潤一・田中 伸佳・石川 雅彦  
(京都大学・工学部)

## 1. はじめに

自然言語の解析において並列句(並列名詞句, 連用中止法による並列動詞句・節)は, 並列の範囲の決定に多くの可能性があり, あいまいさの主たる要因の一つになっている。とくに, 科学技術論文の抄録等では, 少ないスペースに納めべく多くの内容を記述する必要から, 並列表現が多用される傾向にある。本報告では, 実際の科学技術論文にあらわれた並列表現を対象にして, 並列表現の実態調査を行ない, その調査結果をもとにして, 並列表現の解析アルゴリズムを検討した。

## 2. 並列名詞句<sup>(2)(4)</sup>

### 2-1. 並列名詞句の調査

並列名詞句の調査は, 日本科学技術情報センタ(JICST)発行の抄録文(電気工学編Vol. 18 No. 10及びVol. 25 No. 4から783文)に対して行った。ここでいう並列名詞句とは, 表1に示すような助詞・接続詞・句読点によって, 複数個の名詞句が結びれた表現のことである。このように, 複数個の名詞句を結びつける役割をする語(あるいは記号)のことを, 以下では並列接続語とよぶことにする。表1が, 調査データ中にあらわれた並列接続語のすべてである。

並列名詞句が実際にどの程度の解釈のあいまいさをひきおこすかについての調査をまず行った。ここでは名詞句内での係り受け構造のあいまいさおよび意味解釈レベルでのあいまいさは調査の対象とはせず, 並列の範囲決定において生じるあいまいさの教のみを調査している。例えば

(例1) ジャイロトロン内のモード

競合の研究は電子流のジャイロトロピーと動作空間の軸対称性により著しく簡単化できる。

においては並列接続語「と」によって「電子流のジャイロトロピー」と「動作空間の軸対称性」の2名詞句が並列に接続されているが, 意味を考慮に入れなければ種々の組合せが考えられる。すなわち「と」の前方の並列要素としては「ジャイロトロピー」または「電子流のジャイロトロピー」の2通り, また「と」の後方の並列要素としては「動作空間」または「動作空間の軸対称性」の2通りが考えられる。従って例1では図1のように4通りのあいまいさがあることになる。ここでPSYMは並列接続語, P-ELは並列要素, P-BDは並列名詞句全体を表わす。

・ (読点)	・ (中黒)	と	および
または	ならびに	あるいは	や
か	とか	も	もしくは
かつ			

表1 並列接続語

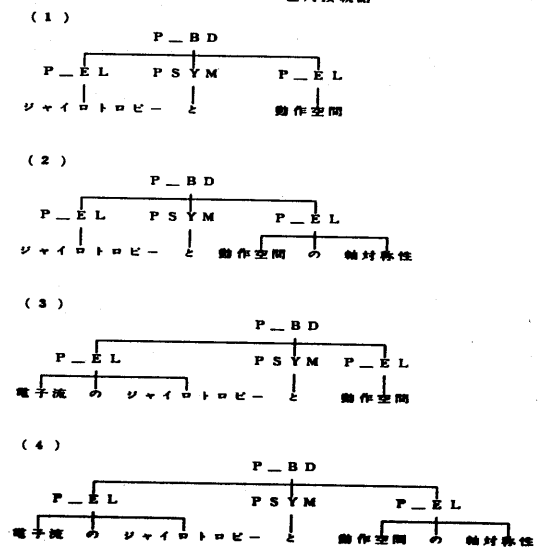


図1 例1の並列名詞句にあらわれるあいまいさ

表2, 表3に一つの並列接続語あたりのあいまいさの数を示す。ただし, あいまいさの数を調べる際, 意味は考慮に入れないが, 非交叉条件などによって統語的に並列要素に成りえないものはあいまいさの数には含めていない。また, 並列接続語が

(例2) 中央研究所の70日プロジェクト「組織構造」投下資本について取り扱った。

(例3) 74年度の米國工業界における動きは, 生産および検査の自動化と電子計算機による管理に移行している。

のように複数個ある場合には, 例2のように同一レベルの並列句を作っている場合と, 例3のようにレベルの異なる並列句が存在する場合とがあるので, 表2, 表3の調査では取扱っていない。これについては2-3で論じる。

あいまいさの数	1	2	3	4	5	6	7	8	9
並列名詞句の数	77	96	41	27	12	3	2	2	0

あいまいさの数	10	11	12以上	計
並列名詞句の数	1	1	0	262

平均のあいまいさの数: 2.40

表2 後方のあいまいさの数の分布

あいまいさの数	1	2	3	4	5	6	7	8	9
並列名詞句の数	145	55	30	14	6	1	0	4	4

あいまいさの数	10	11	12	13	14	15	16以上	計
並列名詞句の数	0	1	1	0	0	1	0	262

平均のあいまいさの数: 2.07

表3 前方のあいまいさの数の分布

がかりとになっているかを示す。

[手がかり1]

強い切れ目を越えて名詞句が並列に接続されることはないとする。すなわち, 強い切れ目を越えて並列の範囲が広がることはないということである。ここで強い切れ目というのは次のようなものである。

(1) 「, (読点)」

(2) 助詞の「は」

ただし, 名詞の直後にある「,」は, それ自身が並列接続語であるので強い切れ目としては扱わない。また, 並列接続語の直後にある「,」は「並列要素の中に含まれている読点」とは考えられないので強い切れ目とはしない。この手がかりはかたまり安定してあり, 強い切れ目を越えて並列の範囲が広がっている例は調査中には存在しなかった。またこれによって, あいまいさの数は表4, 表5のように減少した。

あいまいさの数	1	2	3	4	5	6	7	8
並列名詞句の数	88	105	36	22	8	1	0	1

あいまいさの数	9	10以上	計
並列名詞句の数	1	0	262

平均のあいまいさの数の変化: 2.40 → 2.13

表4 後方のあいまいさの数の変化

あいまいさの数	1	2	3	4	5	6	7	8
並列名詞句の数	146	57	35	14	3	2	2	1

あいまいさの数	9	10	11以上	計
並列名詞句の数	1	1	0	262

平均のあいまいさの数の変化: 2.07 → 1.87

表5 前方のあいまいさの数の変化

## 2-2. 並列名詞句解析の手がかり

並列名詞句の範囲を決定する大きな要因は, 個々の名詞が表れる意味であろう。しかしながら, 機械翻訳等の応用を考える場合には, 個々の名詞ごとに詳細な意味記述が与えられていることはあまり期待できない。そこで, 表層上に明示的にあらわれた手がかりを軸として, どの程度の範囲決定ができるかの調査を行った。以下にそれらの手がかりとそれがどの程度確実な手

[手がかり2]

次のいずれかに該当する「と」は格助詞であると解釈する。

(1) 「と」格をとる動詞・サ変名詞の直前にある「と」。

(2) 「と」格をとる動詞・サ変名詞との間に名詞以外の任意の語が(任意個)存在するような助詞「と」。

この手がかりは並列の範囲限定のため

のものではないが、例えば、

(例4) 夕体問題におけるダイアグラムと類似の説明を行なう。

(例5) 実験値とよく合う事が示された。

のように、並列句とを解釈することが可能であるが実際には並列句にならないものを排除するための手がかりである(表6)。

成功	失敗	計
29	0	29

表6 手がかり2

### [手がかり3]

並列接続語「・」の並列範囲は単一名詞であるとする。この手がかりを適用すれば、

(例6) マイクロプロセッサによる高速な画像データの処理・生成を目指す画像処理装置を…。

では「処理」と「生成」が並列に接続されていることによる(表7)。

成功	失敗	計
28	0	28

表7 手がかり3

### [手がかり4]

助詞「と」と「と」が呼応しているときは、二つの「と」ではさまれた名詞句が後方の並列要素となる。「か」と「か」、「とか」と「とか」についても同様である。この手がかりは、

(例7) ウォルシュ関数のZ変換上インデックスのグレイコード表現との関連を求め…。

のような呼応を扱うためのものである(表8)。

	成功	失敗	計
連体形連の込みを含む	5	0	5
連体形連の込みを含まない	15	0	15
計	20	0	20

表8 手がかり4

### [手がかり5]

並列接続語の後方に「後方の要素を限定する語」が存在する場合、並列接続語の直後の語から「後方の要素を限定する語」に最も近い名詞までを後方の並列要素とする。ここで「後方の要素を限定する語」は次の3種類に分けられる。

- (1) 助詞の「など」
- (2) 並列句を前方に要求しやよい語  
eg. 関係・間・比較・相互作用・比較する

- (3) 次のパターンに現れる「,」

[...N, JT N...]

(N: 名詞, JT: 連体助詞)

これらはそれぞれ次のような並列表現の範囲を決定するための手がかりである(表9)。

(例8) 迷路法, 線分探索法, チャンネル法などの各種配線手法が…。

(例9) 多次元分布のキュミュラントとモーメントの関係と完全に対応してモーメント括弧で表わし, …。

(例10) 理論, 実験, 結果及び考察, の各項について…。

	成功	失敗	計
(1)	24	0	24
(2)	36	2	38
(3)	7	0	7
計	67	2	69

表9 手がかり5

### [手がかり6]

並列接続語に隣接した前後2個の名詞連続について、それらの最初の語同志が等しいか、最後の語同志が等しい場合、その2個の名詞連続が並列に接続されているとする。これは次のような例を対象としたものである(表10)。

(例11) 散乱行列と伝達行列のトポロジ的關係。

(例12) 電子付着と電子離脱の気体過程を…。

成功	失敗	計
62	5	67

表10 手がかり6

[手がかり7]

次のパターンにおいて  $N_1$  と  $N_2$  が同一名詞であるとき、「...  $N_2$ 」の部分を並列の後方の要素とする。

$N_1$  PSYM 「...  $N_2$ 」

手がかり6は、並列接続語に隣接する二つの名詞が共通の語構成要素からできているかどうかを調べるものであったが、この手がかり7は後方の並列要素を修飾句(または節)が修飾している場合を認めるものである。修飾するのが句の場合と節の場合とに分けて調査した(表11)。

	成功	失敗	計
句	13	0	13
節	19	1	20
計	32	1	33

表11 手がかり7

[手がかり8] (表12)

次のいずれかの場合には、前方の連体修飾句が後方の並列要素を修飾することはないとする。

- (1) 並列接続語の直後に「その」など(「そのの」、「それらの」)がある場合。
- (2) 後方の並列要素が連体形埋込みによって修飾されている場合。
- (3) 並列接続語の直前または直後に「(読点)」がある場合。
- (4) 「と」と「と」が呼応している場合(手がかり4参照)。
- (5) 手がかり7が適用できる場合。

[手がかり9] (表13)

前方あるいは後方の並列要素が、何らかの原因によって単一の名詞連続であると判断される場合には、他方の並列要素も単一の名詞連続であるとする。

	成功	失敗	計
(1)	17	0	17
(2)	25	2	27
(3)	20	2	22
(4)	9	0	9
(5)	20	3	23
計	94	7	101

表12 手がかり8

成功	失敗	計
120	29	149

表13 手がかり9

[手がかり10]

連体修飾句を一般に「Xの」という形で示すとすると、「AのBのCのDのEのF」というパターンにおいては、「AのBの(CとD)のEのF」および「(AのBのC)と(DのEのF)」の並列が非常に多い(表13)。したがって、CとDの類似度と、CとFの類似度とを比べてそのいずれかを決定する。

パターン	並列名詞句の数
AのBの(CとD)のEのF	32
(AのBのC)と(DのEのF)	32
その他	18
計	82

表14 手がかり10

この手がかり10は非常に弱い手がかりであり、これでの手がかりで決定できないときに、名詞の細分類や意味的素性を参照することによって構造を決める場合に使用されるべきものである。

2-3. 並列接続語間の優先順位

一つの名詞句の中に二つ以上の並列接続語が存在する場合、それらの接続語が同一レベルの並列を構成しているかどうかを判定する必要がある。例えば、

(例13) 等方性媒質と異方性媒質の境

界、金属と異方性媒質の境界における散乱係数の数値計算結果を…。では、並列接続語「と」と「、」のレベルが異なっている。このように、並列接続語のレベルを正しく認識しないと、非常に多くの解釈の可能性が生じることになる。実際には算術式にあらわれる「+」、「×」等の演算子に優先度があるように、自然言語で使われる並列接続語間にも、並列範囲の広さに関する優先度があると考えられる。ここでは、このような並列接続語間に見られる支配関係について考える。前例のように「と」の才が「、」より強い結合を示し、従って狭い範囲での並列句を構成している場合、

「と」 > 「、」

と表わし、同じレベル内で一つの並列名詞句を構成している場合は

「と」 = 「、」

と表わすことにする。また、ある文の中で最後にある並列接続語にはLASTという特別な語を付記して区別することにすれば、並列接続語間の優先関係は表15のようになる。

また、二つ以上の並列接続語があらわれた場合であって、それらが同一のレベルの並列接続を表わしている場合には、並列接続語の並びに特有のパターンがみられる。出現したパターンを表16に示す。

### 3. 並列名詞句の解析アルゴリズム

2節では、並列名詞句の範囲を決定するのに有効であろうと思われる表層上の各種の手がかりを整理した。これらの手がかりには、必ず成功する非常に確実な手がかりと、失敗する例も多い不確実な手がかりとがある。また、複数個の手がかりが同時に存在し、それぞれが別の範囲決定を示唆する場合もある。このようなことから、確実度の高い範囲決定規則から順に入力文に

並列接続語間の優先関係	頻度
「、」 = 「、」	79
「、」 < 「と」	21
「、」 = 「と (LAST)」	8
「、」 > 「と」	5
「、」 < 「と (LAST)」	5
「、」 < 「および」	6
「、」 = 「および (LAST)」	6
「、」 = 「および」	1
「、」 < 「および (LAST)」	1
「、」 = 「あるいは (LAST)」	2
「、」 = 「や (LAST)」	2
「、」 < 「・」	12
「、」 < 「、」	1
「と」 < 「および」	2
「と」 = 「および (LAST)」	2
「と」 < 「および (LAST)」	2
「と」 > 「および (LAST)」	2
「と」 < 「や」	1
「と」 > 「や」	1
「と」 < 「または (LAST)」	1
「と」 < 「・」	2
「と」 < 「と」	4
「および」 < 「ならびに (LAST)」	1
「および」 < 「および」	1
「とか」 < 「あるいは (LAST)」	1
「・」 = 「および (LAST)」	2
「・」 = 「・」	3
「や」 = 「あるいは (LAST)」	1

(注) 並列接続語の組合せのうち、この表にあがっていないものは頻度が0である。

表15 並列接続語間の優先関係

パターン	頻度
NP, NP, . . . , NP	69
NP, NP, . . . , NP と NP	8
NP, NP, . . . , NP および NP	6
NP, NP, . . . , NP あるいは NP	2
NP, NP, . . . , NP や NP	2
NP と NP および NP	2
NP や NP あるいは NP	1
NP および NP, NP, . . . , NP	1
M · M · . . . · M	3
M · M および M	2

NP: 名詞句 M: 名詞連続

表16 三個以上の名詞句が並列に接続されている場合のパターン

規則を適用してゆき、最後に、範囲決定ができない場合に、例えば手がかり10aのように不確定ではあるが、ほとんどの並列句の範囲を強制的に決定する規則を適用するアルゴリズムが考えられる。現在、我々の開発している文法記述用言語GRADEは、このような順次的処理が可能のように設計されており、このGRADEをアルゴリズムの記述言語として、並列句の範囲決定アルゴリズムを開発している。手がかり1、手がかり2は解析のより前の段階で処理されるとして、それ以降の処理の概略フローを図2に示す。

現在、GRADEのインフォリメンテーションが終った段階であり、ここで示したアルゴリズムを実データに対して適用する実験はまだ行ってないが、思考実験では約85%の並列句の範囲を正しく認定できる。

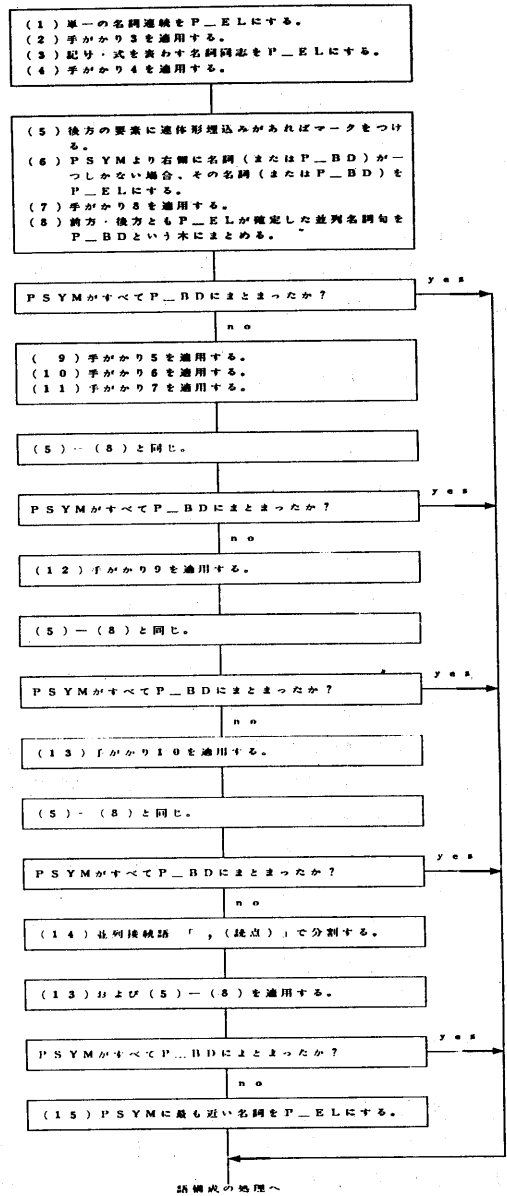
#### 4. 連用中止法の調査

連用中止法は動詞句（あるいは文）の並列を表現する手段であり、名詞句の並列句と同様に日本語の解析を難しくしている問題の一つである。難しくしている原因の一つは、連用中止法が、意味的には様々な関係で結びつく二つの文を、接続助詞の「が」の場合と同様に安易に結びつけてしまうことにある。したがって、並列名詞句の場合と同様に並列範囲の決定に関するあいまいさと同時に、文相互の意味的關係についてのあいまいさがあらわれることになる。

##### 4-1. 連用中止法におけるあいまいさ

###### I. 統語的マスコード

(a) 連用形動詞がそれ以降にあらわれるどの動詞と結びつくのかという点にあいまいさが生じる。例えば次の例では「CADプログラムを利用して少なくさせる」という意味なのか、「C



P-EL : 並列要素  
P-BD : 並列名詞句全体  
PSYM : 並列接続語

図2 並列名詞句解析アルゴリズムの主要部

Aのプログラムを利用して作業を行う」という意味なのかがはっきりしない。

(例14) CADプログラムを利用して論理回路網の変数を少なくさせる作業を行なうことができる。

実際に抄録文中にあらわれた連用中止法がどの程度のあいまいさをひきおこしているかを表17, 表18に示す。表17は調査した文中にいくつの連用形動詞を含んでいるかを示したものであり、表18は連用形動詞を一つだけ含む文についてそれぞれの文のあいまいさの数を示したものである。

連用形動詞数	1	2	3	4
文の数	164	30	3	4

表17 文中の連用形動詞の数

あいまいさの数	1	2	3	4	5	6
文の数	104	44	11	3	1	1

平均のあいまいさの数 : 1.51

表18 連用中止法のあいまいさの数の分布

(b) 連用形動詞の前に連用修飾句があるとき、それが連用中止法によって結びつけられた動詞の両方を修飾するのか、あるいはどちらか一方だけを修飾するのかという問題。

(c) 連用形になった動詞のうち3に付くべき助動詞あるいはアスペクトを示す表現が省略される。文を解析するとき、これらを復元する必要があるかどうかという問題。これについては、「れる」「られる」「させる」等は決して省略できず、「た」は必ず省略される(後の動詞が完了態であれば前の連用形動詞も必ず完了態である)から、実際に問題となるのは「た」に続くタイプの助動詞(「らしい」等)やアスペクトを示す表現(「ている」「ことができる」等)である。

## II. 二文間の関係

前に述べたように、連用中止法に

よって結びつけられる文の間には様々な意味的關係があり、それぞれの意味的關係に従って、連用中止法によって結びつけられる述部間の制約条件も異なってくることを考えられる。ここでは、連用中止法によって結びつけられる文相互の関係を文まかに分類してみた。

A. 連用中止法で結びつけられる二文がほぼ対等に並列要素となっているもの

(i) 二文に分けた方がよいと思われるもの。すなわち並列になっている二文の間にほとんど関係がないもの。並列になり動詞にほとんど制約がない。また埋込み文の中には現れない(26例)。

(例15) 基本的なものは、μPMTランジスタ、ゲート制御ダイオードなどで、装置はすべて学生が操作する。

(ii) 時間的前後関係があるもの(41例)。

(例16) 高電流のパルス放電開始における過渡状態での現象を解析し、実験と比較した。

(iii) 単純な並列。andで結ぶことができる。二文の順序を入れかえても、意味はかわらない(45例)。

(例17) 負荷は純抵抗性であって整合回路で使用される素子のQは、あらかじめ定められた値を越えない。

(iv) 二文が逆接の関係にあるもの(2例)。

(例18) 第一の実験ではTownsend放電からの全放射の測定を行ない、それによって、放電における光電現象に対する充満値説明ができた。

B. 連用中止形の節が後にくる節の連用修飾節になっていると考えられるもの

(i) 内容状況を示すもの。「～したから」「～という状態(状況)で」と言いかえられるもの(7例)。

(例19) ベル研究所はウエスタンエレクトリック社と協力して、各種軍用機密の研究・開発および製造につとめてきた。

(ii) 手段を示すもの。「～することによって」と言いかえられるもの。前にくる動詞としては「使う」「用いる」等が多い。また、後にくる動詞が道具(手段)格をとる必要があると考えられる(20例)。

(例20) ここでは線材の断面を円形として計算し、ケーブルとしての損失上昇係数を求めた。

(iii) 目的を示すもの。「～する下めに」と言いかえられるもの(1例)。この用例は、「～に応じて」といった固定的な表現をとる場合に限られており、一種の慣用句処理とするか、あるいは格助詞相当表現と考える処理するのが妥当だと思われる。

(iv) 前提条件を示すもの。「～した上で」と言いかえられるもの。前にくる動詞としては「仮定する」「考慮する」等がある(10例)。

(v) 原因・理由を示すもの。「～したので」と言いかえられるもの(5例)。

(例21) 二次放射線の放射波はスフェロイド波となりこのため場は速度依存性をもち。

(vi) 結果を示すもの。「～した結果」と言いかえられるもの。後にくる動詞としては「得る」等の特別な動詞が多い(4例)。

(例22) また、マイクロ波干渉法と探針測定の結果とを比較し、フラズマ

のイオン組成の情報も得た。

B. の下位分類の多くは連用中止形になる動詞にかなり強い制限があり、各動詞に個別的情報を付け加えておくことにより処理すべきであると思われる。

#### 4-2. 動詞の分類と連用中止法

連用中止法を結びつけられる二文のうち、特にA.の分類を結びつけられる二文の述部間には、かなりの類似度が見られるはずである。ここでは、この類似度の尺度として、動詞を他動詞・自動詞、意志動詞・無意志動詞、動作動詞・状態動詞に分類<sup>(5)</sup>し、連用中止法を結びつけられる二つの動詞の共起関係を調査した。結果を表19に示す。この表から同一カテゴリに分類される動詞同志の共起が著しいことがわかる。例えば意志動詞・無意志動詞の組合せを禁止するという規則を設ければ、この規則によって連用中止法のスコアのあいまいさは大きく減少することになる(表20)。

前の動詞	後の動詞	文の数
自動詞	自動詞	19
自動詞	他動詞	14
他動詞	自動詞	7
他動詞	他動詞	123

前の動詞	後の動詞	文の数
動作動詞	動作動詞	133
動作動詞	状態動詞	4
状態動詞	動作動詞	9
状態動詞	状態動詞	17

前の動詞	後の動詞	文の数
意志動詞	意志動詞	119
意志動詞	無意志動詞	6
無意志動詞	意志動詞	7
無意志動詞	無意志動詞	31

表19 結びつけられる動詞のカテゴリの組合せ

あいまいさの数	1	2	3	4	5	6
適用前	0	44	111	3	1	1
適用後	24	28	4	2	1	0

表20 規則適用によるあいまいさの数の減少



#### 4-3. 連用中止法に関する

##### 表層上の手がかり

4-2. で述べた動詞の分類を用いる手法は、並列名詞句の範囲決定に名詞意味分類を用いるものに対応しており、このような語の類似性による以前に、表層上の手がかりを用いることも可能である。そのような手がかりとしては以下のものが考えられる。

(i) 連用形の動詞と非連用形の動詞が連続していたら（誤点があってもいい）、それらは並列になっている。

(例23) 形状の影響を実際に研究し計算するのに十分足るような……。

(ii) 同一の動詞は並列になっていることが多い（しかもその関係は4-1. で述べたII-(iii) (単純な並列になっている)。

(例24) 無限区間を有し、振動する被積分項を有する収束のゆるい積分の数種の評価を論ずる。

(iii) 接続助詞「が」「と」、接続助詞的名詞「場合」「ため」「とき」「結果」等を越えて連用中止になる並列の範囲が広がることはない。

(例25) 長波尾の波形を印加した時、電離波が平板から戻り、リーダーチャンネルと遭遇すると、絶縁破壊が起る事が示された。

(iv) 論文抄録特有の動詞（「述べる」「検討する」「紹介する」「示す」「記述する」「論ずる」「指摘する」など）は最高レベルに現われる。

(例26) ここではその欠点を指摘し、この改良法としてマイクロ流システムに普及している2ポート回路による解析法を提示した。

現在、このような様々な手がかりを用いた連用中止法の範囲決定アルゴリズム

を作成中である。

#### 5. まとめ

並列構造は自然言語解析における解釈のあいまいさの主要な要因の一つになっている。本稿では、並列の範囲を決定する際に、単なる品詞ならびのような統語的な構造だけでなく、様々な表層表現上の手がかりが存在することに注目し、それらの手がかりがどの程度有効であるかを、実際のデータを調査することにより明らかにした。これらの手がかりを規則として定式化し、思考実験を行うことにより、深い意味解析を行わずともかなりの高精度で並列句の範囲を決定できることを示した。

実際に文法記述用言語 GRADE でこれらの規則を書き、実験することは今後の課題であるが、従来の統語処理・意味処理の難しさを抑えたい、解析上の各種の手がかりがあるということは重要である。

#### 参考文献

- 1) 佐藤理史, 「機械翻訳のための文法記述言語 GRADE の作成」, 京都大学工学部特別研究報告書 (1983)
- 2) 千田裕彦, 「文法記述言語 GRADE におけるパターンマッチングプログラム」, 京都大学工学部特別研究報告書 (1983)
- 3) 茅賀純, 「現代日本語の文法」, (教育出版, 1978)
- 4) 久野暁, 「日本文法研究」, (大修館書店, 1973), p. 68 ~ p. 75
- 5) 金田一春彦編, 「日本語動詞のアスペクト」, (みぎ書房, 1976)