

## 日本語文における名詞句の並列構造の推定 および推敲支援への適用

山村 広臣   菅沼 明   牛島 和夫  
九州大学 工学部 情報工学科

並列構造を含む文は、どの文節列とどの文節列が並列しているかという並列要素の認識において曖昧性を含みやすい。また、文章推敲の立場から考えても、文の書き手と読み手の間に食い違いが生じやすい。本研究は、文章推敲の立場から名詞句の並列構造を推定し、書き手と読み手の間に食い違いが生じやすい名詞句の並列構造を指摘することを目的としている。名詞句の並列構造を推定する際、並列要素を決定する表層的な手がかり、および並列要素の構造的類似性に着目して、推定を行なった。JICSTの抄録文に対して実験を行なったところ、文章中の並列のキーのうち96%を抽出できた、また、抽出した並列のキーのうち69.7%は、並列要素を正しく推定できた。これによって上記の着目点が有効に働いていることが分かった。また、本手法を推敲支援へ適用した。

### Analysis of Coordinate Structures in a Japanese Document, and its Application to a Writing Tool

Hiroomi YAMAMURA, Akira SUGANUMA, Kazuo USHIJIMA  
Department of Computer Science and Communication Engineering  
Kyushu University  
6-10-1 Hakozaki, Higashi-ku, Fukuoka, 812 Japan

In Japanese documents, sentences with coordinate structures are often ambiguous when two elements of the coordinate structure are recognized. It is often the case that the reader cannot recognize the elements of the coordinate structure in a sentence which the writer intends. Then, some communication gap occurs between the writer and the reader. In this paper, we describe the method to estimate coordinate structures consisting of noun phrases. Our method analyzes a Japanese document with only textual information. The method estimates the elements of the coordinate structure with a clue of Japanese characters and structural similarity between two elements. We made an experiment to the coordinate structures in JICST abstracts with the method. The precision, that is the proportion of the correctly estimated structures versus the extracted ones, of the method is 69.7%. We apply this method to a writing tool.

## 1. はじめに

近年、機械翻訳や文章校正支援に代表される自然言語処理技術の発展は着実である。しかし、いまだ困難な問題はいくつも残されている。その中の一つに並列構造の解析がある。並列構造とは、一文中に同等の機能を持つ複数個の文節列を並べた構造である。並列構造を含む文は、どの文節列とどの文節列が並列しているかという並列要素の認識において曖昧性を含みやすい。そのため並列構造を誤って認識してしまうことが多く、後の構文解析や意味解析にも解析誤りが生じてしまう。また、並列構造を含む文を文章推敲の立場から考えても、文の書き手と読み手の間に食い違いが生じやすい。

ここで、読み手がどのようにして並列要素を判断しているか考えると、1) 文全体を見て、大まかな構文を把握する、2) 並列構造の存在を示す語の前後の文節の意味的な類似性や構文的な類似性を発見する、が挙げられる。これらを機械処理で実現するには、すべての構文の可能性とすべての単語の意味を蓄積するための莫大な記憶領域と、多くの解析時間が必要になる。

並列構造を推定するには、名詞の意味情報やそれに伴う制約から解析することが一般的である<sup>[1, 2]</sup>。しかし、同じような意味情報を持つ名詞が複数存在する場合は、必ずしも正しい並列構造を決定できるとは限らない。また、文章推敲という立場から考えると、分かりにくい並列構造を正確に推定できなくても、十分に使用価値は存在する。つまり、書き手と読み手の間に食い違いが生じやすい並列構造を指摘することが要求される。

本研究は、文章推敲の立場から名詞句の並列構造(名詞並列と呼ぶ)を推定し、書き手と読み手の間に食い違いが生じやすい名詞並列を指摘することを目的としている。本研究では名詞並列を推定する際、構造的類似性に基づいて推定する。また、並列構造の推定を推敲支援に適用するために、ユーザに煩わしさを感じさせない待ち時間で処理をしたいという要求がある。そこで、大規模な辞書の使用、単語の意味を反映した解析、形態素解析を行っていない。このことにより並列構造を推定する精度は落ちるが、素早い処理が期待できる。

## 2. 名詞並列の推定

本研究では、名詞並列のみを対象としている。名詞並列の存在を示す語を並列のキーと呼び、その前後の並列要素を前置要素、後置要素と呼ぶことにする。また、並列要素の最初と最後の文節を起点文節、終点文節と呼ぶことにする(図1)。並列のキーとしては、「[読点], [中点], と, も, や, 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」をとりあげている。

### 2.1 処理の概要

読み手が文章を読む時、並列構造の曖昧性に直面することがよくある。このようなときは意味的に類似もしくは対比している単語に注目して並列要素を決定することが多い。しかし、このような単語が複数存在したり、存在しなかったりする場合もある。このような場合でも、構造的に類似している名詞並列は読み手に正確に伝わると考えられる。また、実際に、名詞並列の前後の文節列は構造的に類似していることが多い。そこで本研究では、構造的類似性に基づいて名詞並列を推定する。構造的類似性に基づく推定とは、例えば、「. . AのBとCのD. . .」という名詞並列があれば、(AのB)を前置要素、(CのD)を後置要素とする処理を行なうことを意味する。もし、この文の(B)と(C)が実際に並列していたとしても意味的に類似もしくは対比している名詞がない場合、読み手は一意に並列要素を決定できないであろう。

名詞並列を推定する手順は、以下のとおりである。

1. 主に字種情報を利用して、文を仮の文節に分割する(2.2.1節)。
2. 仮の文節に文節の性質を付与し(2.2.2節)、文節間の係り受け関係を決定する(2.2.4節)。ただし、文節の性質や文節間の係り受け関係は、並列要素の構造的類似性を比較するための情報である。
3. 主に構造的類似性に基づいて並列要素を決定する(2.3節)。

### 2.2 前処理

構造的類似性に基づいて名詞並列を推定する際、名詞並列の前置要素と後置要素の候補の構造を比較する。比較する情報として、文節の性

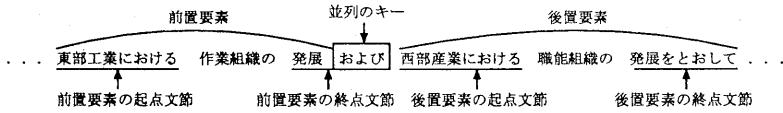


図 1: 名詞並列の各部名称

質と文節間の係り受け関係を用いる。前処理として、仮の文節切りを行ない、これらの情報を得る。できるだけ素早い処理で、これらの情報を抽出することが目的である。

### 2.2.1 仮の文節切り

日本語文は、漢字仮名交じりのべた書きで記述されるのが普通である。漢字仮名交じりの日本語文の特徴として、自立語、特に数的にも多い名詞や、動詞や形容詞の語幹は漢字で表記され、逆に付属語や副詞や接続詞などは平仮名書きされ易い。そこで、主に字種情報を利用することにする<sup>1</sup>。平仮名から非平仮名への字種の変わり目で仮の文節<sup>2</sup>に分割する。しかし、字種情報を利用する場合には、(i) 分割不足、(ii) 分割位置誤り、(iii) 過剰分割という分割誤りが生じてしまう。図 2 は字種情報による分割誤りの例である。

お | 父さんは | いつも | 走り | 回っている  
(ii) (i) (iii)

図 2: 字種情報による分割誤りの例

これらの分割誤りについては、文節の性質の決定を経て 2.2.3 節で修正を行なう。

### 2.2.2 文節の性質の決定

文節の性質とは、文節が持ち得る資格である。文節の性質は、文節に係り得る資格と受け得る資格から成る。この二つの資格をそれぞれ係りの資格、受けの資格と呼ぶことにする。係りの資格と受けの資格の分類の方法は難しい問題であるが、文献[4, 5, 6]を基に表層表現で得られるように分類した。表 1 は、係りの資格の分類である。表 2 は、受け資格の分類と、受けの資格が受けることが可能な係りの資格の対応表である。係りの資格は文節を構成する最後の単語によって決定し、受けの資格は文節の自立語によって決定する。

<sup>1</sup>字種情報のみで 84% の精度で正しく分かち書きが行なわれたという報告もある<sup>[3]</sup>。

<sup>2</sup>仮の文節は、1 個以上の自立語 + 0 個以上の付属語から成る単語列と定義する。以下では、単に文節と呼ぶ。

表 1: 係りの資格の分類

| 分類              | 説明            |
|-----------------|---------------|
| a: 格要素 / が格     | 格助詞「が」        |
| b: 格要素 / を格     | 格助詞「を」        |
| c: 格要素 / に格     | 格助詞「に」        |
| d: 格要素 / で格     | 格助詞「で」        |
| e: 格要素 / から格    | 格助詞「から」       |
| f: 格要素 / と格     | 格助詞「と」        |
| g: 格要素 / より格    | 格助詞「より」       |
| h: 格要素 / へ格     | 格助詞「へ」        |
| i: 連体修飾要素 / 連体詞 | 連体詞による連体修飾    |
| j: 連体修飾要素 / 用言  | 用言による連体修飾     |
| k: 連体修飾要素 / の格  | 格助詞「の」による連体修飾 |
| l: 連用修飾要素 / 副詞  | 副詞による連用修飾     |
| m: 連用修飾要素 / 用言  | 用言による連用修飾     |
| n: 接続要素         | 接続助詞などによる接続関係 |
| o: 主題要素         | とりたてて「は」      |
| p: 並列のキー要素      | 名詞並列の存在を示す表現  |
| q: 文末           | 文末            |

表 2: 受けの資格の分類

| 分類            | 受けることが可能な係りの資格                     |
|---------------|------------------------------------|
| 名詞            | i, j, k                            |
| 動詞の連体形        | a, b, c, d, e, f, g, h, l, m, n    |
| その他の動詞        | a, b, c, d, e, f, g, h, l, m, n, o |
| 形容詞の連体形       | a, c, g, k, l                      |
| 形容動詞の連体形      | a, c, d, g, k, l                   |
| 形容詞, 形容動詞の連用形 | l                                  |
| その他の形容詞, 形容動詞 | a, c, d, e, f, g, h, l, m, n, o    |
| 判定詞の連体形       | a, c, d, e, g, i, j, k, l, m, n    |
| その他の判定詞       | a, c, d, e, g, i, j, k, l, m, n, o |
| 受けなし          |                                    |

文節の性質を決定するために、我々の研究室で考案された活用チェック法<sup>[7]</sup>を利用する。活用チェック法を適用すれば、活用語の品詞やその活用形の推定が可能である。本研究では、文節の最後の文字から文節の先頭に向かって文字列の接続を調査することによって、文節の性質を決定する。文節の性質を決定する手順は、以下のとおりである。

I: 文節の最後の文字に活用チェック法を適用する。連体形と一致すれば、係りの資格を「連体修飾要素 / 用言」とする。連用形と一致すれば、係りの性質を「連用修飾要素

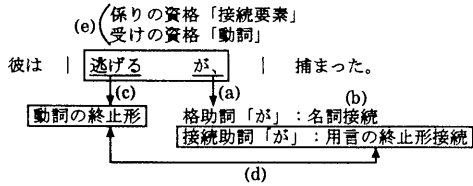


図 3: 文節の性質を決定する例

/用言」とする。受けの資格は活用チェック法で推定された品詞により決定する。

- II: 文節の最後の文字列に助詞をマッチさせる (図 3-(a)). マッチした助詞が接続する品詞の候補を得る (図 3-(b)). その接続候補に対して、接続条件を満たしているかどうかを調査する (図 3-(d))<sup>3</sup>. 接続条件を満たしていたら、受けの資格を実際に接続している品詞 (および活用形) によって決定する。係りの資格はマッチした助詞によって決定する (図 3-(e)).

例えば、図 3は助詞「が」(格助詞と接続助詞の2つが考えられる)と接続する自立語を推定することにより、文節「逃げるが、」の係り資格と受けの資格を決定している。

### 2.2.3 分割誤りの修正

本節では、2.2.1節での文節の分割誤りを修正する。

**分割不足の修正** 分割不足の主な原因は、文節(自立語)が平仮名で始まる文節を分割できなかったことにある。この問題点を補うために、次の二つの処理を追加する。

- 平仮名書きされやすい自立語を登録し、その自立語に字面でマッチすれば文節とする<sup>4</sup>。
- 特定の文字(格助詞と並列のキー)に対する状態遷移表<sup>5</sup>を用いて、文節に分割する。

**分割位置誤りの修正** 文節の性質を決定できなかった文節は分割位置誤りによって生成された文節とし、その文節と後の文節をつなげて一つの文節とする。

**過剰分割の修正** 過剰分割の主な原因は、複合語を分割してしまったことにある。そこで、分

<sup>3</sup> 接続候補が用言である場合は、活用チェック法を用いて調査する (図 3-(c)).

<sup>4</sup> 平仮名書きされやすい自立語として、副詞と接続詞の約 1200 単語を登録したテーブルを使用している。

<sup>5</sup> これは、特定の文字が文節の切れ目になるかを決定するための状態遷移表である。約 2000 万字の文章を基に、人手で決定した。

表 3: 複合語の処理

|    | 前側の文節 + 後側の文節 | → 新しい文節 | 例      |         |
|----|---------------|---------|--------|---------|
| 係り | 連用/用言         | *       | *      | 飲み+続ける  |
| 受け | 動詞            | 動詞      | 動詞     | → 飲み続ける |
| 係り | 連用/用言         | *       | *      | 聞き+難い   |
| 受け | 動詞            | 形容詞     | 動詞、形容詞 | → 聞き難い  |
| 係り | 連用/用言         | *       | *      | 降り+口    |
| 受け | 動詞            | 名詞      | 名詞     | → 降り口   |

割された複合語を復元することにより修正を行なう。前後の文節の係りの資格と受けの資格を基に、表 3の条件に当てはまる前後の文節は複合語とし、一つの文節にする。また、複合助詞についても一つの文節にする (例: に対して)。

### 2.2.4 文節間の係り受け関係の決定

係り受け解析の大きな問題として、並列構造の認定がある。この問題に対処する方法として、はじめに並列構造を決定してしまう方法や、並列構造の存在を考慮しながら係り先を決定していく方法がある [6, 8]。しかし本研究では、並列のキーを含む文節以外の係り受け解析を先に行ない、その後で並列要素を決定する手法をとる。並列構造の範囲を絞り、並列要素を決定するための情報として、文節間の係り受け関係を利用する。

係り先を決定する際、素早い処理で実現するために、単語の意味情報まで考慮せず、文末から順に係り先を一つに決定していく。係り受け解析の手順は以下のとおりである。

- I: 係りの資格と受けの資格を基に係り先の候補を求める (表 2)。
- II: 一般則と経験則を用いて係り先を一つに絞る。一般則とは、例外が稀であり、強い規制力をもつ規則である。表 4に一般則と経験則を示す。

表 4: 係り受け関係を決定する一般則と経験則

| 一般則                                     |
|---|
| 各文節はそれより文末側にある一つの文節に係り得る                |
| 各文節はそれより文頭側にある 0 個以上の文節を受け得る            |
| 係り受け関係は互いに交差しない                         |
| 経験則                                     |
| いくつかの例外 <sup>6</sup> を除いて、係り得る最も近い文節に係る |
| 読点を伴う文節は連体修飾文節には係り得ない                   |

<sup>6</sup> 例外処理として、文献[8]にある読点や助詞「は」の処理などがある。

### 2.3 並列要素の推定

名詞並列の構造的類似性に基づいて並列要素を決定する。具体的には、前置要素と後置要素の構造木をつくり、その構造木の構造を比較して最も類似している要素を並列要素と推定する。しかし、この処理だけでは正しい並列要素を推定することは困難である。そこで、以下の手順で並列要素を推定する。ただし、これは一文中に並列のキーが1つある場合の手順である。複数ある場合については2.3.4節で説明する。

- (1) 名詞並列の最長範囲を決定する(2.3.1節)。
- (2) 最長範囲内で並列要素を決定する表層的な手がかりを用いて、並列要素を推定する(2.3.2節)。
- (3) 前置要素と後置要素の比較対象となる構造木を作り、構造的類似性に基づいて並列要素を推定する(2.3.3節)。

#### 2.3.1 名詞並列の最長範囲の決定

すべての前置要素と後置要素の候補を比較することは、多くの間違った候補の生成や並列要素の推定誤りを引き起こす。そこで、名詞並列の最長範囲を決定する必要がある。これにより間違った並列要素の候補の数を減らすことができる。

名詞並列の前方最長範囲を、並列のキーを含む文節から文頭の文節に向かって、2.2.4節で得られた文節間の係り受け関係を基に決定する。前方最長範囲の決定に関して、対象文節が一つ後の文節から並列のキーを含む文節までに係る場合は最長範囲内であるとする。また、名詞並列の後方最長範囲を、並列のキーの後の文節から文末の文節に向かって決定する。後方最長範囲の決定に関して、対象文節が連体修飾要素または並列のキーを含む文節に係る、または並列のキーを含む文節や連体修飾要素である場合は最長範囲内であるとする。

図4-(a)の場合、上記の処理を行なうことで、前方最長範囲は文頭までとなり、後方最長範囲は文節「人が」までとなっている。図4-(b)の場合、文節「彼女は、」が文節列「太郎が友達から借りたノート、」に係っていないので、名詞並列の前方最長範囲は文節「太郎が」までとなっている。また、文節「鉛筆を」が文末の文

節に係っているので、名詞並列の後方最長範囲は文節「鉛筆を」までとなる。

#### 2.3.2 並列要素を決定する表層的な手がかり

並列構造を含む文には、読み手が並列要素を容易に決定できる表現がよくある。このような表現を見つけ出すことは、並列要素を決定する手がかりとなる。本研究では、以下の4つを手がかりとして利用している。

助詞の共起 助詞が並列のキーと共起して並列要素を明らかにする表現である。

(例) 評価の必要性と投資の問題とを。 . . .

後置表現 いくつか列挙した事柄を総括する、もしくはそれによって他にも同類の事柄が存在することを暗示する表現である。一般的には並列句の直後に置かれるので有用な手がかりとなり得る。

(例) 方法の改善や加工の合理化などに. . .

東ねの用法 名詞句を並列に並べることによって新しい概念を表示する用法である。特定の表現と共起している名詞並列は東ねの用法と解釈される場合が多い。

(例) 一つの入力と出力から成る. . .

形態的類似性 並列要素の起点文節や終点文節が互いに同じ漢字やカタカナ列を含む場合は並列要素を決定する表層的な手がかりになり得ると考えられる。また、同じ漢字を含むことは意味的に何らかの類似性があるはずである。

(例) 必要な信頼性、保守性、改善性を. . .

本研究では、これらの手がかりを登録したテーブルを使用している。名詞並列の最長範囲内で字面で手がかりにマッチする文節が存在する場合は一意に並列要素を決定する。

#### 2.3.3 構造的類似性に基づく並列要素の推定

名詞並列の前置要素と後置要素に並列要素を決定する表層的な手がかりが存在しない場合には、構造的類似性に基づいて並列要素を決定する。その際、名詞並列の前置要素と後置要素になり得る候補を比較し、構造的に最も類似している要素を並列させる。そのために、まず前置要素と後置要素を比較するための構造木(比較対象構造木<sup>7</sup>と呼ぶ)をつくる。これは最

<sup>7</sup>文節をノード、文節間の係り受け関係をブランチとする構造木である。

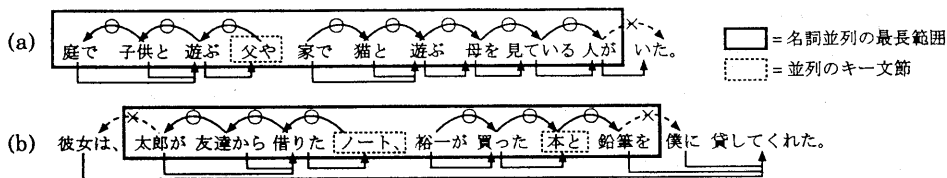


図 4: 名詞並列の最長範囲の推定例 (a) 並列のキーが1つ (b) 並列のキーが複数の場合

長範囲内の並列要素の候補を得るための処理でもある。前置要素の比較対象構造木は、並列のキー文節をルートノードとし、前方最長範囲までの文節を順に付け加えた構造木である。後置要素の比較対象構造木は、受けの資格が「名詞または判定詞」である文節をルートノードとし、並列のキーを含む文節の後の文節からルートノードの文節までをノードとして持つ構造木である。ただし、受けの資格が「名詞または判定詞」であっても、その文節を飛び越えて他の文節に係る文節が存在する場合はルートノードとしない。図 4-(a)における比較対象構造木を図 5に示す。文節「家で」が文節「猫と」を飛び越えて他の文節に係るので、文節「家で猫と」が後置要素の候補になり得ない。

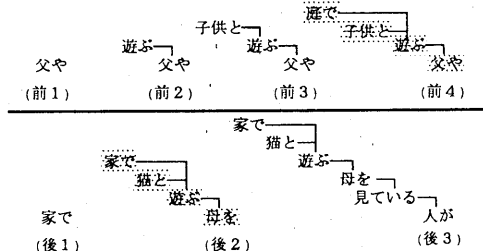


図 5: 図 4-(a)における比較対象構造木

次に、並列要素を一つに推定する処理を行なう。構造的類似性に基づいて並列要素を推定する。その情報として、今までの処理で得られた文節の性質と文節間の係り受け関係を利用する。前置要素と後置要素の比較対象構造木をルートノードから末端ノードまで比較し、以下の優先順位で並列要素を一つに決定する。

- 1: 比較対象構造木が完全に一致する要素を並列させる。
- 2: 比較対象構造木が部分的に一致する要素を並列させる。このとき、できるだけ多くの文節が一致する要素を並列させる。
- 3: 並列要素の最長範囲を並列させる。

図 5において、比較対象構造木(前 4)と(後 2)が完全に一致するので、文節列「庭で子供と遊ぶ父や」が前置要素、「家で猫と遊ぶ母を」が

後置要素となる。

しかし、以下に示す特定の場合には、経験則を用いて並列要素を決定する。

- a. 前置要素の最長範囲内の文節数が1つ、または後置要素の最長範囲内の文節数が1つしかない場合、並列要素を決定する表層的な手がかりがなければ、以下の規則を適用する。

| 名詞句の構造    | 適用規則             |
|-----------|------------------|
| A の B と C | B を前置要素, C を後置要素 |
| A と B の C | A を前置要素, B を後置要素 |

- b. 並列のキーが特殊キー<sup>8</sup>である場合、並列要素は最長範囲とする。並列要素が長くなる場合は書き手が意図的もしくは無意識に特殊キーを使うため、この経験則を用いる。
- c. 後置要素が連体詞「その、この」で始まる場合、その連体詞が長い前置要素を指し示すことが多いので、前置要素は最長範囲とする。
- d. 並列要素の終点文節が形式名詞を含む場合、その形式名詞が長い連体修飾要素を受けることが多いので、並列要素は最長範囲とする。

### 2.3.4 並列のキーを複数含む名詞並列の推定

名詞並列の最長範囲内に複数の並列のキーを含む場合は、各並列のキーに対する並列要素を推定し、最後に各並列構造間の関係を整理することにより推定を行なっている。本研究では、並列構造は他の並列構造と交差しないということ为前提にしている。その理由は、並列構造が交差していると読み手に読みにくく、また誤った並列構造を推定することが多くなるからである。推定の手順は、以下のとおりである。

手順 1: 並列要素を推定する順番を並列のキーのレベルにより決定する。並列のキーを「読点以外」の並列のキー(レベル 1)、「読点」の並

<sup>8</sup> 「読点を除く並列のキー+読点」または「読点+読点を除く並列のキー」となっている並列のキー。(例)「と、」、「または、」

列のキー(レベル2), 特殊キー(レベル3), の三つのレベルに分割し, レベル1に相当する並列のキーから順に並列要素を推定していく。

図4-(b)において, 並列のキー「(ノート)、」はレベル2, 並列のキー「(本) と」はレベル1に相当するので, 並列のキーを含む文節「本と」について先に並列要素を推定していく。

手順2: 各並列のキーに対する最長範囲を決定し, 構造的類似性に基づいて並列要素を推定する。ただし, 並列のキーに対する最長範囲の推定方法は, 基本的には2.3.1節と同じであるが, 並列要素を決定していない並列のキーを含む文節を越える範囲を指定できないとする。

図4-(b)において, 並列のキー文節「本と」の最長範囲は「裕一が買った本と鉛筆を」になる。また, 2.3.3節の経験則 a. より前置要素が「本と」, 後置要素「鉛筆を」に推定される。次に, 並列のキーを含む文節「ノート、」の並列要素を推定する。最長範囲は, 並列のキーを含む文節「本と」の並列要素が決定しているので, 「太郎が～ノート、～鉛筆を」になる。比較対象構造木をつくると図6のようになる。(前4)と(後2)が部分的に3文節一致するので, 「太郎が～ノート、」が前置要素, 「裕一が～鉛筆を」が後置要素になる。

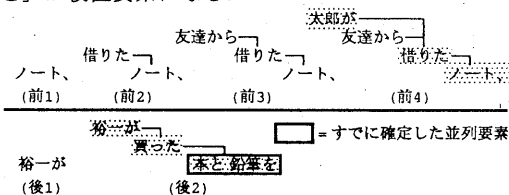


図6: 図4-(b)における比較対象構造木

手順3: 並列構造間の関係を整理する。本研究では並列構造の交差は認めていないので, 並列構造間の関係は兄弟関係と親子関係と他人関係だけである。兄弟関係とは, 2つの並列構造がまったく同じ並列要素をもつ場合である。親子関係とは, 一方の並列構造に他方の並列構造が含まれる場合である。他人関係とは, 2つの並列構造の並列要素が全く重ならない場合である。例えば, 図4-(b)の並列構造間の関係は親子関係になる(図7)。

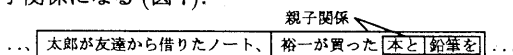


図7: 図4-(b)における並列構造の推定結果

### 3. 文章への適用

本手法を計算機上に実装し, JICSTの抄録文(299件, 54,858文字)に対して名詞並列の自動推定を行なった。推定結果の判別については人手で行なった。表5は, 各並列のキーに対する並列要素の推定結果を, 並列のキーを1つもつ名詞並列と複数もつ名詞並列に分けて, 示した表である。また, 実際の並列要素の文節数に対する推定精度も示している。文章を入力し並列構造が得られるまでの解析時間は, SPARC station ELC(CPU: SPARC/33MHz)上で, 1万字の文章に対して約1.3秒程度であった。

並列のキーを1つもつ名詞並列の推定精度は75.4%(正解: 325個, 不正解: 106個)であった。並列のキーを複数もつ名詞並列の推定精度<sup>9</sup>は59.3%(正解: 99個, 不正解: 68個)であった。また, 並列のキーとして適当でないものが38個あり, 並列のキーとして抽出できなかったものが37個(文章中の並列のキーのうち約4%)あった。

誤った並列要素を推定した原因を, 並列のキーを1つもつ名詞並列と複数もつ名詞並列のそれぞれについて, 表6に示す。

誤った並列要素を推定した原因の上位3つは原因A, B, Cである。そこで, これらの原因について考察を行なう。2.3.3節の経験則 a. より正しい並列要素を決定できたものは324個(正解率: 86.4%)であり, 例えば, 「AのBとC」となる文はBとCを並列させることが有効である。また, 構造の完全一致または部分一致により正しい並列要素が決定できたものは121個(正解率: 71.6%)あり, 構造的類似性に注目して名詞並列を推定する方法は間違っていないと考えられる。また, 並列要素を決定する表層的な手がかりにより, 正しい前置要素または後置要素を決定できたものは533個とかなり多くあった。

本手法における, 並列のキーを複数もつ名詞並列の推定精度は, あまりよくない。主な原因の一つは, 並列要素の推定順序を誤ったことにある。そのため, まったく異なった並列構造間の関係が得られてしまった。もう一つは, 並列

<sup>9</sup>すべての並列のキーに対して, 正しい並列要素が推定できている場合を正解とする。

表 5: 名詞並列の推定結果

| 実際の並列要素の文節数        | 並列のキーを1つもつ名詞並列 |      |      |      |      | 並列のキーを複数もつ名詞並列 |      |      |      |      | 計    |
|--------------------|----------------|------|------|------|------|----------------|------|------|------|------|------|
|                    | 1              | 2    | 3    | 4    | 5以上  | 1              | 2    | 3    | 4    | 5以上  |      |
| 正しい並列要素を推定できたもの    | 253            | 37   | 14   | 7    | 14   | 280            | 64   | 15   | 8    | 10   | 702  |
| 誤った並列要素を推定したもの     | 33             | 30   | 15   | 13   | 15   | 51             | 48   | 26   | 13   | 23   | 267  |
| 並列のキーとして適当でないもの    |                |      |      |      |      |                |      |      |      |      | 38   |
| 正解率(%)             | 88.5           | 55.2 | 48.3 | 35.0 | 48.3 | 84.6           | 57.1 | 36.6 | 34.8 | 30.3 | 69.7 |
| 並列のキーとして抽出できなかったもの |                |      |      |      |      |                |      |      |      |      | 37   |

のキーがすべて読点である場合の推定誤りにある。並列のキーがすべて読点である場合、並列構造間の関係はすべて兄弟関係になることが多いので、特別な処理をしてもよいと考えられる。また、名詞並列の中に述語並列が入れ子として存在することもあるので、これも考慮する必要がある。

#### 4. 推敲支援への適用

並列構造を含む文を文章推敲の立場から考えると、並列構造の範囲やその前後の文節間の係り受け関係に曖昧性を含むことが多いので、読み手と書き手の間に食い違いが生じやすい。本節では、並列構造の推定を推敲支援に適用する方法について述べる。本研究では、1) 問題となりそうな並列構造があればそれを提示する、2) 本手法を用いて得られる情報を提示する、ことを推敲支援への適用方針とした。

1) に関して、書き手と読み手の解釈に食い違いが生じやすい並列構造を一般的に規定するのは難しい。しかし、並列要素を決定する表層的な手がかりが存在する並列構造を分かりやすい並列構造とし、それ以外の並列構造を提示することにした。2) に関して、以下の2つの推敲支援を行なっている。

- 本手法を用いて並列要素の最尤候補、または部分一致候補、または全候補<sup>10</sup>のいずれかを書き手に提示する。
- 並列構造の問題点を提示する。  
(例) 並列要素の最長範囲が長い。

#### 5. おわりに

本研究では、大規模な辞書の使用、単語の意味を反映した解析、形態素解析を行わずに、名詞並列の推定を行なった。その際に、構造的類似性に基づく推定方法について述べた。今後

<sup>10</sup> 最尤候補：本手法による解析結果、部分一致候補：構造が部分的に一致する候補、全候補：本手法を用いて得られた全ての候補

表 6: 誤った並列要素を推定した原因

- (1): 並列のキーを1つもつ名詞並列  
(2): 並列のキーを複数もつ名詞並列

| 原因                      | (1) | (2) | 計   |
|-------------------------|-----|-----|-----|
| A: 2.3.3 節の経験則 a. による誤り | 26  | 25  | 51  |
| B: 構造の完全・部分一致による誤り      | 23  | 25  | 48  |
| C: 表層的な手がかりによる誤り        | 19  | 20  | 39  |
| D: 前処理の誤り               | 13  | 9   | 22  |
| E: 並列構造が部分並列であった        | 7   | 5   | 12  |
| F: 名詞並列の最長範囲の推定誤り       | 5   | 4   | 9   |
| G: 並列要素の推定順序の誤り         |     | 15  | 15  |
| H: 以前の並列要素の推定誤りの影響      |     | 38  | 38  |
| その他                     | 13  | 20  | 33  |
| 計                       | 106 | 161 | 267 |

の課題として、部分並列の考慮、述語並列への拡張が挙げられる。

#### 参考文献

- [1] 田村, 田中: 意味解析に基づく並列名詞句の構造解析, 情報処理学会自然言語処理研究会, 59-2, (1987)
- [2] 黒橋, 長尾: 長い日本語文における並列構造の推定, 情報処理学会論文誌, Vol.33, No.8, pp.1022-1031(1992).
- [3] 坂本: 文節の認定, 日本語情報処理シンポジウム報告集, 情報処理学会プログラミング・シンポジウム委員会 (1978)
- [4] 吉田: 二文節間の係り受けを基礎とした日本語文の構文分析, 電子情報通信学会誌, Vol.55-D No.4, pp.238-244 (1972)
- [5] 首藤, 楳原, 吉田: 日本語の機械処理のための文節構造モデル, 電子情報通信学会誌, Vol.62-D No.12, pp.872-879 (1979)
- [6] 山上, 安原: 形態素情報による日本語の係り受け解析, 情報処理学会自然言語処理研究会報告, 88-1 (1992)
- [7] 笠原, 菅沼, 牛島: 日本語文章推敲支援ツール『推敲』における活用チェック法による否定表現の抽出, 第 51 回情報処理学会全国大会論文集, 2H-2 (1995)
- [8] 黒橋, 長尾: 並列構造の検出に基づく長い日本語文の構文解析, 情報処理学会自然言語処理研究会報告, 98-2 (1993)