日本語文における名詞句の並列構造の推定およびその推敲支援への適用

菅沼 明[†] 山村広臣[††,☆] 牛島和夫[†††]

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，文章の書き手と読み手の間に食い違いが生じやすい．したがって，文章を推敲する観点から考えても，並列構造は無視できない．本研究は，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並列構造を文章推敲の立場から指摘することを目的としている．本稿で述べる並列構造の推定法は，ユーザに煩わしさを感じさせない待ち時間で処理を行うために，字面の情報を主に利用する．表層的な情報を利用して文節に区切り，文節の性質と係り受け関係を推定する．それを用いて並列要素を推定する．並列要素の推定では，並列要素を決定する表層的な手がかり，構造的類似性および経験則に基づいて推定を行う．本手法の評価として，JICST の抄録文（約 5 万字）を解析対象の文章として，並列構造の推定実験を行った．その結果，文章中に存在する並列のキーのうち 97.3 %を抽出できた．抽出した並列のキーのうち 70.5 %は，並列要素を正しく推定できた．また，SPARCstation ELC を用いて処理時間を測定した結果，上記の抄録文中のすべての並列構造を推定するのに約 7.1 秒を要した．さらに，本手法を推敲支援へ適用するプロトタイプを作成した．

Analysis of Coordinate Structures in Japanese Documents, and Its Application to a Writing Tool

AKIRA SUGANUMA,[†] HIROOMI YAMAMURA[††,☆] and KAZUO USIIIJIMA[†††]

In Japanese documents, sentences with some coordinate structures are often ambiguous when two components of the coordinate structure are recognized. It is often the case that the reader cannot recognize the coordinated components which the writer intends. As a result, some communication gaps occur between the writer and the reader. The writer must be careful about coordinate structures, so it is a useful function in writing to indicate the presence of coordinate structures in a sentence. In this paper, firstly, we describe the method for extracting coordinate structures consisting of noun phrases and estimating the coordinated components. Our method analyzes a Japanese document with only textual information, so that the writer can comfortably find information of coordinate structures in a few seconds. The method estimates the components using clues on Japanese characters, heuristics to determine the coordinate structure, and the structural similarity between two components. Secondly, we evaluate our method with JICST abstracts (about 50 thousand Japanese characters). This method extracts candidates of coordinate structures from the abstracts and estimates the coordinated components of each candidate. Consequently, 97.3 % of all the coordinate structures in the abstracts are extracted by this method. The precision, that is the proportion of the correctly estimated structures versus the extracted ones, is 70.5 %. Finally, we apply the method to a writing tool. We have implemented a prototype with this method.

† 九州大学大学院システム情報科学研究科知能システム学専攻
Department of Intelligent Systems, Kyushu University
†† 九州大学大学院工学研究科情報工学専攻
Department of Computer Science and Communication Engineering, Kyushu University
††† 九州大学大学院システム情報科学研究科情報工学専攻
Department of Computer Science and Communication Engineering, Kyushu University
☆ 現在，NTT データ通信株式会社
Presently with NTT Data Communications Systems Co.

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は着実に発展している．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 1 つに並列構造の解析がある．並列構造とは，1 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

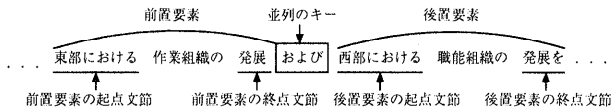


図1 並列構造の各部名称
Fig. 1 An example of the coordinate structure.

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を生じてしまう．また，文章を推敲する観点から考えても，書き手と読み手の間で並列構造の解釈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，をあげることができる．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，多くの解析時間を必要とする．

並列構造を推定するには，単語の意味情報やそれにともなう制約から解析することが一般的である[1),2)]．しかし，同じような意味を持つ単語が存在しなかったり，複数存在する場合には，必ずしも正しい並列構造を決定できるとは限らない．一方，文章推敲の立場から考えると，分かりにくい並列構造を正確に推定できなくても，並列構造の存在の可能性を指摘できればツールとしての使用価値は十分ある．

本研究は，日本語文章中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並列構造を指摘することを目的としている．文章の推敲支援に応用することを考えると，処理時間が長ければ書き手の思考を妨げることになる．そのため，ユーザに煩わしさを感じさせない待ち時間で処理したいという要求がある．この要求を満たすために本研究では，大規模な辞書を使わず，単語の意味を反映した解析や形態素解析を行わない手法を採用する．これによって，並列構造を推定する精度は必ずしも高くないが，素早い処理が期待できる．

2. 並列構造

2.1 並列構造に関する用語

文中に存在する並列構造は，たとえば，**図1** のような構造を持つ．並列構造の存在を示す語を**並列のキー**と呼ぶ．並列のキーを挟んで並立する2つの文節列を**並列要素**と呼び，前側を**前置要素**，後側を**後置要素**と呼ぶ．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．図1にあるように前置要素の終点文節は並列のキーの文節であり，後置要素の起点文節は並列のキーの文節の直後の文節となる．

表1 並列のキー
Table 1 Keys to coordinate structures.

名詞並列	[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，及び，又は，並びに，或は
部分的並列	[読点]，だけで（は）なく，および，または，ならびに，あるいは，もしくは，及び，又は，並びに，或は

1文中に複数の並列のキーが現れる場合，読み手が並列構造を把握できるものとして，一方の並列構造が別の並列要素に完全に含まれる場合（**親子関係**），2つの並列構造がまったく同じ並列要素を共有する場合（**兄弟関係**），2つの並列構造の並列要素がまったく重ならない場合（**他人関係**）の3通りが考えられる．これらのうち親子関係と兄弟関係は，2つの並列構造が重なりあっているため，**複合並列構造**と分類する．また，並列のキーが1文中に1つしか現れない並列構造と他人関係を**単一並列構造**と分類する．

本研究では，文献3) に倣い，並列構造を (a) 文形式の並列構造（**述語並列**と呼ぶ），(b) 名詞句の並列構造（**名詞並列**と呼ぶ），(c) 非名詞句の並列構造（**部分的並列**と呼ぶ）の3つに分類する．ここで，部分的並列とは，文形式から右端の部分語列を述語込みで取り去った残りの語列で，単一の名詞句と見なせないものが並列要素となる並列構造を指す．述語並列，名詞並列，部分的並列の例を以下にあげる．下の例の『』で囲んだ部分が並列要素である．

(a) 待ち合わせにおいて，『ロック要求が拒否され』，かつ『デッドロック要求が検出されない』とき ...
(b) 業務を『排他性の原則』と『構造性の原則』によって配分し ...
(c)『人口は10倍』，『情報利用は40倍』に ...

本稿では，名詞並列と部分的並列を並列構造の推定対象としている．また，並列のキーとして，**表1** に示したものを取り上げる．

2.2 推定処理の概要

読み手が文章を読むとき，並列構造の曖昧性に直面

することがよくある．このような場合には，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．この類似もしくは対比する単語は並列構造中に存在しない場合もあれば，複数存在する場合もある．その場合，読み手と書き手の間に食い違いが生じることが多い．しかし，類似もしくは対比する単語が手がかりにならない場合でも構造的に類似している並列構造であれば，読み手に正確に伝わることがある．これは，読み手が構造的類似性を手がかりにして並列要素を判断することを意味している．また，実際に，並列構造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並列構造を推定する．この方法をとることで読み手の並列構造の認識に近付ける．構造的類似性に基づく推定とは，たとえば，「...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)を後置要素とすることを意味する．しかし，並列構造を判断する単語が明示されている場合には，読み手はその単語をもとに並列要素を決定する．このことを反映するため，並列要素を決定する表層的な手がかりおよび経験則を利用する．

名詞並列と部分的並列の並列構造を推定する手順は，前処理と本処理の2つに分けることができる．それぞれ以下の手順で処理を行う．

前処理：

(1) 主に字種情報を利用して，文を仮の文節に分割する（3.1節）．

(2) 仮の文節に文節の性質を付与し（3.2節），それを用いて仮の文節の分割誤りを修正する（3.3節）．

(3) 文節の性質を利用して文節間の係り受け関係を決定する（3.4節）．

本処理： 前処理で求めた文節の性質と文節間の係り受け関係を用いて並列構造の推定を行う．

(1) 文章中の並列のキーを検索し，各並列のキーの前後の文節列に対して並列構造になりうる最長の文節列を求める（4.1節）．

(2) 構造的類似性に基づいて部分的並列の並列要素を推定する（4.2節）．

(3) (2)で並列要素を決定できなかったものに対して，名詞並列の並列要素を推定する（4.3節）．

(4) (1)で求めた最長の文節列内に複数の並列のキーを含む並列構造の推定を行う（4.4節）．

3. 前処理

3.1 仮の文節切り

本研究では，大規模な辞書を使わず，形態素解析も行わずに，日本語文の表記の特徴に着目して文節切りを行う．日本語文は，一般に漢字仮名交じりのべた書きで記述される．漢字仮名交じりの日本語文の特徴として，動詞・形容詞の語幹や名詞は漢字で表記され，逆に付属語や副詞や接続詞などは平仮名書きされやすい．そこで，平仮名から非平仮名への字種の変わり目で仮の文節に分割する．仮の文節は，1個以上の自立語＋0個以上の付属語からなる単語列と定義する．以下では，仮の文節を単に文節と呼ぶ．

字種情報を利用して文節に区切る場合に起こる誤りの修正は，3.2節で述べる文節の性質を利用して行うものがあるので，3.3節で説明する．

3.2 文節の性質の決定

文節の性質は，文節が係りうる資格（**係りの資格**）と受けうる資格（**受けの資格**）からなる．これらの資格をどのように分類するかは非常に難しい問題である．本研究では，文献4)〜6)で提案されたものをもとに，表層表現で判断できるように分類した．**表2**は，係りの資格の分類表であり，**表3**は，受けの資格の分類と，受けることが可能な係りの資格の対応表である．なお，これらの分類は，あくまで表層的な構造を表すものであり，深層的な意味を表すものではない．

本研究では，文節を構成する最後の単語によって係りの資格を決定し，文節の自立語によって受けの資格を決定する．この処理を行うために，我々の研究室で

表2 係りの資格の分類
Table 2 Classification of the property of modifying phrase.

分類	説明
a：格要素/ガ格	格助詞「が」
b：格要素/ヲ格	格助詞「を」
c：格要素/ニ格	格助詞「に」
d：格要素/デ格	格助詞「で」
e：格要素/カラ格	格助詞「から」
f：格要素/ト格	格助詞「と」
g：格要素/ヨリ格	格助詞「より」
h：格要素/ヘ格	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/ノ格	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	とりたて詞「は」
p：並列のキー要素	並列構造の存在を示す表現
q：文末	文末

表3 受けの資格の分類
Table 3 Classification of the property of modified phrase.

分類	受け可能な係りの資格
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	

考案した活用チェック法[7]を利用する．活用チェック法は，部分的な文字列に対して，それが活用語の一部であるかを調べる手法であり，活用語の品詞と活用形の推定が可能である．活用チェック法は，語幹と活用語尾に関する約32 KBの表しか使用しない．また，部分的な文字列しか検査しないので，活用形の推定が素早く行える．このため，受けの資格を決定する方法として活用チェック法を使用する．

本研究では，文節の最後から先頭に遡って文節内の文字列の接続を検査することで，文節の性質を決定する．その手順を以下に示す．

手順I： 文節の最後が用言であると仮定して文節の性質を決定する．文節の最後の文字から活用チェック法を適用する．用言の連体形もしくは連用形と推定された場合は，係りの資格を「連体修飾要素/用言」(表2, j)，「連用修飾要素/用言」(表2, m) とする．受けの資格は，活用チェック法で推定された品詞と活用形により決定する．ただし，文末の場合は用言の終止形であるかを検査し，終止形の場合は活用チェック法で推定された品詞と活用形で受けの資格を決定する．それ以外は受けの資格を「名詞」とする．

手順II： 自立語 + 助詞と仮定して文節の性質を決定する．

(a) 文節の最後の文字と助詞を照合する．

(b) (a) の結果，照合する助詞があれば，その助詞の前側に接続可能な品詞の候補と活用形を得る．

(c) 助詞が用言と接続可能であれば，助詞の1文字前から活用チェック法を適用する．用言であると推定された場合，自立語は活用チェック法で推定された品詞とする．用言でないと推定された場合，自立語は名詞とする．

(d) (c) で推定した自立語が (a) で得られた助詞と接続可能であれば，文節の性質を決定する．係りの資格は，助詞が持つ係りの資格によって決定する．受けの資格は自立語の品詞と活用形によって決定する．

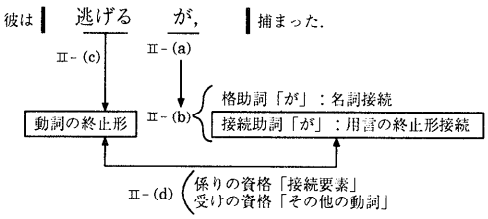


図2 文節の性質を決定する例
Fig. 2 Process to decide the property of phrase with textual analysis.

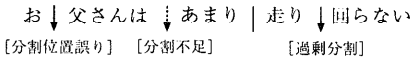


図3 字種情報による分割誤りの例
Fig. 3 An example of incorrect separating with textual information.

たとえば，**図2**は，助詞「が」(格助詞または接続助詞) と接続する自立語を推定することにより，文節「逃げるが,」の文節の性質を決定している．

3.3 分割誤りの修正

字種情報のみで分割した文節の分割誤りには，**図3**に示すように分割不足，分割位置誤り，過剰分割の3種類が考えられる．これら分割誤りのうち分割不足は文節の性質を決定する (3.2節) 前に修正している．

(1) 分割不足の修正

分割不足とは，文節の切れ目となるところを分割できなかった誤りを指す．その主な原因は，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，以下の2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たテーブル☆を利用する．登録した自立語に字面で照合する文字列が存在すれば，それを文節とする．
- 特定の文字 (格助詞と並列のキー) が文節の切れ目になるかを判定するために，状態遷移表を利用する．特定の文字の2文字後の平仮名までを検査し，決定論的に特定の文字が文節の切れ目になるかを決定する．たとえば，図3において，「は」の後に「あ」がつながる自立語は存在しないので，「父さんは」と「あまり」の2つの文節に分割する．

(2) 分割位置誤りの修正

分割位置誤りとは，文節の切れ目でないところを分割した誤りを指す．この誤りが起こると文節の性質を

☆ 平仮名書きされやすい自立語として，副詞，接続詞，連体詞の約1200単語を登録したテーブルである．

決定できない場合が多い．そのため，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した文節とし，その文節と後の文節をつなげて1つの文節にする．たとえば，図3において，文節「お」の文節の性質を決定できないので，後の文節「父さんは」とつなげて1つの文節「お父さんは」にする．

(3) 過剰分割の修正

過剰分割とは，1つの文節を2つ以上の文節に分割してしまった誤りを指す．その主な原因は，複合語を分割してしまったことにある．そこで，分割された複合語を復元することで修正を行う．連続する文節の係りの資格と受けの資格を検査し，**表4**の条件に当てはまる連続する文節を複合語と見なして，1つの文節にする．また，複合助詞（「に対して」など）についても1つの文節にする．

3.4 文節間の係り受け関係の決定

文節間の係り受け関係を求める方法として，はじめに並列構造を決定してしまう方法[8]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[6]がある．しかし，本研究では，並列のキーの文節以外の係り受け解析を先に行い，係り受け関係を利用して並列要素を推定する．

文節間の係り受け関係を素早く決定するために，日本語の特徴と表層的な表現とを利用する．**表5**に示した一般則と経験則を用いて，文末の文節から文頭の文節に向かって順に係り先を一意に決める．

手順I： 係り先を求める文節の係りの資格と各文節の受けの資格とを比較して，表3を満たす係り先の候補を求める．係り先を求める文節より文末側の文節に対してこの処理を行う．

手順II： 一般則の非交差条件と経験則を用いて係り先を1つに絞る．手順Iで求めた係り先の候補について非交差条件を満たすものだけを候補として残す．この段階で1つに絞られていなければ，経験則を満たすものを選択する．

表4 複合語の処理
Table 4 Process for a compound word.

	前側の文節 ＋ 後側の文節 → 新しい文節			例
係りの資格	連用/用言	*	#	飲み ＋ 続ける
受けの資格	動詞	動詞	動詞	→ 飲み続ける
係りの資格	連用/用言	*	#	聞き ＋ 難い
受けの資格	動詞	形容詞	動詞，形容詞	→ 聞き難い
係りの資格	連用/用言	*	#	降り ＋ 口
受けの資格	動詞	名詞	名詞	→ 降り口

*：どんな係りの資格でも可
#：後側の文節の係りの資格

表5 係り受け関係を決定する一般則と経験則
Table 5 Rules to determine the relation of the phrases.

一般則	各文節はそれより文末側にある1つの文節に係りうる
	各文節はそれより文頭側にある0個以上の文節を受けうる
	係り受け関係は互いに交差しない（非交差条件）
経験則	いくつかの例外を除いて，係りうる最も近い文節に係る
	読点をともなう文節は連体修飾文節には係りえない

4. 並列要素の推定

並列構造の構造的類似性に基づいて名詞並列と部分的並列の並列要素を推定する．この推定を行う際には，前処理で得られた「文節の性質」と「文節間の係り受け関係」を用いる．前処理が終了した段階では，並列のキーの存在は分かるが，名詞並列と部分的並列の区別はできていない．そのため，並列構造の最長範囲の推定（4.1節）を経て，部分的並列の並列要素の推定（4.2節），名詞並列の並列要素の推定（4.3節）の順で並列要素の推定を行う．

4.1 並列構造の最長範囲の推定

並列構造の並列要素を推定する際，並列のキーの前後で並列要素となりうる候補をでたらめに求めても，処理の無駄である．また，多くの間違った候補の生成は，並列要素の推定誤りを引き起こす．したがって，並列要素となりうる最長の文節列を求めることが必要になる．この最長の文節列の範囲を，**並列構造の最長範囲**と呼ぶ．

並列構造の最長範囲を求めるには，前置要素となりうる最長の文節列の起点文節と，後置要素となりうる最長の文節列の終点文節，を求めればよい．これらの範囲をそれぞれ**前方最長範囲**，**後方最長範囲**と呼ぶ．ただし，並列構造の最長範囲はあくまで最長の候補であり，実際の並列要素とは限らない．

図4に並列構造の最長範囲の推定例を示す．図中の四角で囲まれた部分がそれぞれ名詞並列，部分的並列の最長範囲と推定される文節列である．文の下側に記している矢印は，文節の係り先を示している．また，文の上側に，各文節が名詞並列の最長範囲の判定条件を満たすか否かを記している．

4.1.1 名詞並列の最長範囲

名詞並列の前方最長範囲を求めることは，前置要素の終点文節に係りうる最長の連体修飾成分を求めることである．並列のキーの文節から文頭に遡って検査し，検査対象文節が**表6**に示す判定条件を満たす場合は前方最長範囲内であるとする．条件を満たさない文節が出現したら，その文節の直後の文節までを前方最長範囲とする．

後方最長範囲を求めることは，並列のキーに対する

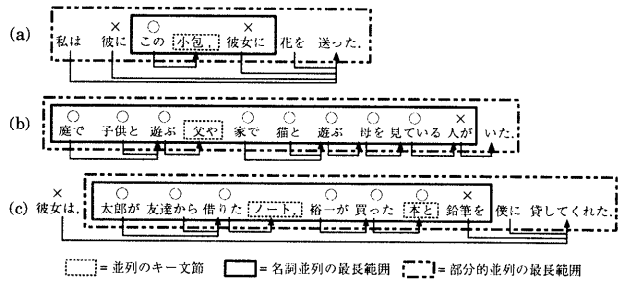


図4 名詞並列と部分的並列の最長範囲の推定例
Fig. 4 Examples of estimating the longest range of coordinate structure.

表6 名詞並列の最長範囲を求めるための判定条件
Table 6 Conditions to determine the longest candidate for the noun phrase.

前方	検査対象文節が，その直後の文節から並列のキー文節までのいずれかの文節に係る
後方	以下の2つの条件のうち，いずれかを満たす ○ 検査対象文節の係りの資格が「連体修飾要素」または「並列のキー要素」である ○ 検査対象文節が，「連体修飾要素」の文節または「並列のキー要素」の文節に係る

後側の最長の名詞句を求めることである．並列のキーの文節の直後の文節から文末に向かって検査し，検査対象文節が表6に示す判定条件を満たす場合は後方最長範囲内であるとする．条件を満たさない文節が出現したら，その文節までを後方最長範囲とする．

最長範囲の推定は文頭に近い並列のキーから順に行う．そのため，複数の並列のキーが存在する文では，後方最長範囲を求める際に別の並列のキーが含まれる場合がある．その場合は複合並列構造と見なし，4.4節で述べる方法で処理する．

たとえば，図4の例文(a)では，文節「彼に」が文節列「この小包,」のいずれの文節にも係らないので，名詞並列の前方最長範囲は文節「この」までになる．並列のキーの後側では，文節「彼女に」が文末に係る連用修飾要素で条件を満たさないので，名詞並列の後方最長範囲は「彼女に」までとなる．また，例文(b)と(c)についても上記の処理を行うと，図に示した範囲が名詞並列の最長範囲となる．例文(c)の場合は，並列のキーの文節「ノート,」の後方最長範囲に並列のキーの文節「本と」が存在するので，複合並列構造と見なして，4.4節で述べる処理に移る．

4.1.2 部分的並列の最長範囲

名詞並列の最長範囲は，名詞句となりうる最長の文節列を決定すればよかった．しかし，部分的並列に関しては，どのような文節列も並列要素となりうる．そこで，意味的な区切りになる文節までを最長範囲とする．表層表現だけで，意味的な区切りになることが分かるものには読点がある．本手法では，部分的並列の最長範囲の推定において読点を区切りとしている．前方最長範囲を求める際には，並列のキーの文節から文頭に向かって1文節ずつ検査し，並列のキーでない読点が出現したら，その文節の直後の文節までを最長範囲とする．読点を含む文節が出現せずに文頭の文節に到達した場合には，文頭までを最長範囲とする．次に，後方最長範囲を求める際には，並列のキーの文節から文末に向かって1文節ずつ検査し，並列のキー以外の読点を含む文節か文末の文節かのどちらかに到達すれば，そこまでを最長範囲とする．

4.2 部分的並列の並列要素の推定

部分的並列は並列要素を対比させる意味で使用することが多い．また，その前置要素と後置要素は構造的に完全に一致することが多い☆．このことから，部分的並列を構造的な一致のみで決定する．

まず，部分的並列の最長範囲内で，部分的並列の前置要素と後置要素になりうる候補を作成する．部分的並列は非名詞句を並列要素として持つので，2つ以上の名詞句からなる文節列が並列要素の候補になる．前置要素の起点文節が名詞並列の最長範囲内にあると前置要素が1つの名詞句となるので，部分的並列の前置要素は名詞並列の最長範囲より前にある文節から始まるものを候補とする．

一方，後置要素の候補は，名詞並列の最長範囲内にも非名詞句になる文節列が存在する．たとえば，図4の例文(b)の文節列「家で猫と」がそれである．そこで，文節の受けの資格が「名詞または判定詞☆☆」で

☆ 文献3)では，部分的並列は並列要素の文節数が一致し，その文法・意味機能が等しい文節が並列していると報告している．また，技術雑誌（「日経コンピュータ」誌などの解説文約22万語）を調査した結果，対称性が認められない部分的並列は，243個の部分的並列のうち十数例しか見られなかったという報告もある．

☆☆「だ」「である」などの断定の助動詞で終わる文節を指す．

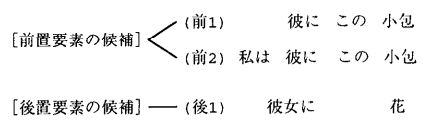


図5 部分的並列の並列要素の候補
Fig. 5 To estimate the candidate of the coordinate structure of the non-noun phrases.

あり，かつ，後置要素の起点文節とその文節との間に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，その文節を後置要素の終点文節の候補とする．**図5**は，図4の例文(a)に対する部分的並列の並列要素の候補を示している．

次に，前置要素と後置要素の候補を比較し，以下の手順で部分的並列の並列要素を推定する．この処理で並列要素が決定できない場合は，4.3節の名詞並列の推定に処理を移す．

手順I： 候補内の各文節の「文節の性質」が完全に一致する候補があれば，それらを部分的並列の並列要素とする．

手順II： 候補内の格要素の構成が完全に一致する候補があれば，それらを部分的並列の並列要素とする．

図5の前置要素と後置要素の候補を比較すると，(前1)と(後1)の格要素の構成がともに「AにB」の形になっているので部分的並列と推定する．また，図4の例文(b)と(c)では，前置要素の候補がないので部分的並列ではないとする．

4.3 名詞並列の並列要素の推定

4.3.1 表層的な手がかりに基づく並列要素の推定

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層的な表現がある．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有用な手がかりとなる．これを**表層的な手がかり**と呼ぶ．本研究では，以下の4つを表層的な手がかりとして利用する．例において，四角で囲んだ部分が表層的な手がかりで並列要素と決定する部分である．

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 | 重要な | 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，それによってほかにも同類の事柄が存在することを暗示する表現である．一般に並列句の直後に置かれるので有用な手がかりとなる．後置表現には，「など，などといった，などという，といった，という」がある．

[例] 方法の | 改善や | 加工の | 合理化などが

束ねの用法 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[3]．特定の表現と共起している名詞並列は束ねの用法と解釈される場合が多い．

[例] 1つの | 入力と | 出力から | 成る

意図的表現 意図的表現とは，書き手が意図的に読み手に並列要素を示している表現である．意図的表現には，読点で並列要素の終わりを示す「読点＋助詞」がある．「受けの資格」が名詞であり，かつ意図的表現を含む文節が存在する場合は，一意に後置要素を決定する．

[例] 効果のある | 手段としての | 組織設計と，| その | 判定基準の | 解析，の | 必要性を

本手法では，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並列のキーの文節から文末に向かって1文節ずつ検査して，テーブルの手がかりが最初に出現した文節までを並列要素とする．

4.3.2 構造的類似性に基づく並列要素の推定

並列要素を決定する表層的な手がかりで並列要素を決定できなかった場合には，構造的類似性に基づいて並列要素を決定する．その際に，まず前置要素と後置要素になりうる候補をそれぞれ作成し，それらを比較して構造的に最も類似している要素を並列要素とする．

前置要素の候補は，並列のキーの文節を終点文節とし，名詞並列の前方最長範囲内の文節を順に付け加えた文節列である．後置要素の候補は，受けの資格が「名詞または判定詞」である後方最長範囲内の文節を終点文節として持つ文節列である．ただし，並列のキーの文節の直後の文節とその文節との間に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，その文節を終点文節の候補としない．

たとえば，図4の例文(b)に対する名詞並列の候補を**図6**に示す．文節列「家で猫と」は，文節「家で」が文節「猫と」を飛び越えて他の文節に係るので非名詞句であり，後置要素の候補にならない．

構造的類似性を利用して並列要素を推定する際には，まず以下の経験則に適合する並列要素を決定する．

[経験則a] 前方最長範囲内の文節数が1であるか，または後方最長範囲内の文節数が1である場合，**表7**の規則を適用する．

[経験則b] 並列のキーが「読点以外の並列のキー＋読点」または「読点＋読点以外の並列のキー」（**特殊キー**と呼ぶ）である場合，並列要素が長くなることが多いので，前置要素と後置要素は名詞並列の最

長範囲とする．

[経験則 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，前置要素は最長候補とする．

[経験則 d] 前置要素または後置要素の終点文節が形式名詞で終わる場合，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，前置要素と後置要素は最長候補とする．

上に示した経験則を用いても並列要素が決定できない場合には，前置要素の候補と後置要素の候補とを順に比較し，並列要素を1つに決定する．処理の手順は以下のとおりである．

手順 I： 候補内の各文節の「文節の性質」と「文節の係り受け関係」を比較し，完全に一致する候補が存在すれば，それらを並列要素とする．

手順 II： 候補内の格要素の構成を比較し，完全に一致する候補が存在すれば，それらを並列要素とする．

手順 III： 候補内の格要素の構成を比較し，部分的に最も多く一致する候補を並列要素とする（**部分一致候補**）．ただし，2文節以上が部分的に一致する候補だけを対象とする．

手順 IV： 前置要素と後置要素の候補の中で，文節数が最も少ない候補を並列要素とする．

図6では，(前4)と(後2)の各文節の「文節の性質」と「文節の係り受け関係」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．

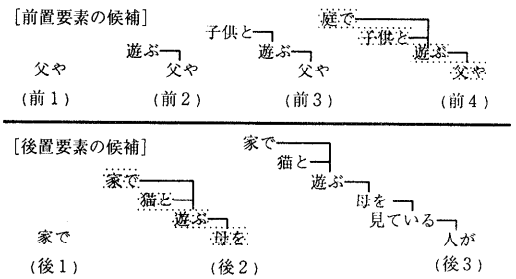


図6 名詞並列の並列要素の候補

Fig. 6 To estimate the candidate of the coordinate structure of the noun phrases.

表7 経験則 a

Table 7 Heuristics for finding a short coordinate structure.

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

4.4 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，以下の手順で並列要素の推定を行う．

手順 I： 並列要素を推定する順序を**並列のキーのレベル**により決定する．並列のキーを中点（レベル1），[読点]以外の並列のキー（レベル2），[読点]（レベル3），特殊キー（レベル4）の4つのレベルに分割し，番号が大きいほどレベルが高いものとする．並列要素を推定する順序はレベルが低い並列のキーから順に並列要素を推定する．

手順 II： 各並列のキーに対する名詞並列の最長範囲を求め，名詞並列の並列要素を推定する．ただし，名詞並列の最長範囲の推定方法は，基本的には4.1節で述べた方法と同じであるが，並列要素を決定していない並列のキーの文節を越えないとする．また，構造の類似性を判定する際には，すでに決定している並列構造の文節列は1文節として扱う．

手順 III： 後から推定した並列のキーの前置要素または後置要素が，すでに決定している並列構造の文節列のみの場合は，これらの並列構造の関係が兄弟関係となるように並列要素を修正する．

たとえば，図4の例文(c)において，並列のキー「ノート，」はレベル3，並列のキー「本と」はレベル2であるので，並列のキー「本と」に対する並列要素を先に推定する．並列のキー「本と」に対する名詞並列の前方最長範囲は，並列のキー「ノート，」の並列構造が求まっていないので，「裕一が」の文節までとなる．並列要素の推定は，4.3.2項の[経験則 a]により，前置要素が「本と」，後置要素が「鉛筆を」となる．次に，並列のキー「ノート，」に対する名詞並列の並列構造の推定を行う．名詞並列の最長範囲は「太郎が～ノート，～鉛筆を」になる．前置要素と後置要素の候補を作成すると，**図7**のようになり，(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」が後置要素と推定する．

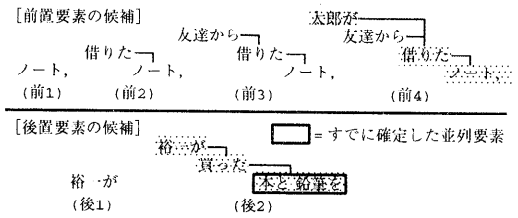


図7 名詞並列の並列要素の候補（並列のキーが複数存在する場合）

Fig. 7 To estimate the candidate of the coordinate structure of the noun phrases (including coordinate keys in the longest range of coordinate structure).

上に示した処理で推定を行うと，並列構造間の関係は親子関係，兄弟関係，他人関係のいずれかになる（2.1 節参照）．図 4 の例文 (c) では，並列構造間の関係は親子関係になる．

5. 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）に対して名詞並列と部分的並列の並列構造の自動推定を行った．推定結果の判別は人手で行った．**表 8** に，並列のキーの抽出結果を示す．

本手法では，形態素解析などを行わずに，字面だけの情報で並列のキーを文章中から抽出している．そのために，並列のキーを見落としたり（第一種の誤り），並列のキーとしては適当でないものを抽出したり（第二種の誤り）する可能性がある．上記の抄録文を用いた評価の際には，第一種の誤りを 27 個，第二種の誤りを 49 個犯していた．抄録文に含まれる並列のキーが 1,006 個で，本手法で指摘した並列のキーが 1,028 個であったので，適合率は 95.2％であり，再現率は 97.3％であることになる．

第一種の誤りの原因として出現頻度が多かったものは，「動詞の転成名詞 +［読点］」の表記で並列のキーとなる場合であった．動詞の転成名詞は動詞の連用形と同じ表記をする．そのため，活用チェック法で動詞の連用形 +［読点］と推定されてしまい，名詞並列の並列のキーとして抽出できない．このために第一種の誤りとなる．

第二種の誤りの原因として出現頻度が多かったものは，「副詞 +［読点］」や「ト格」があった．これらに関しては以下のように考える．

表 8 並列のキーの抽出結果
Table 8 Extraction result of the coordinate key with our method.

抽出した並列のキー		1,028	
	正しい並列のキー	979	(95.2％)
	キーでないものをキーと判定	49	(4.8％)
文中の並列のキー		1,006	
	抽出できた並列のキー	979	(97.3％)
	キーの見逃し	27	(2.7％)

- 「副詞 +［読点］」の例を以下にあげる．
  「昭和 46 年 2 月以来，総合オンライン・システムが稼働している」
  この表現に関しては，辞書を使用して副詞の存在を明らかにすることも考えられる．しかし，副詞と同じ表現で名詞となる場合も存在するので，その場合には第一種の誤りとなってしまう．また，上に示した例では，文意を得ることができなければ，並列のキーと区別することはできない．
- ト格を必須格として要求する用言を登録し，その用言に係る文字「と」で終わる文節をト格と判断することは可能である．形態素解析・構文解析などで正確に係り受けが解析できれば，この第二種の誤りに関しては出現数をさらに減らすことができるであろう．

抽出した**並列のキー** 1,028 個に対して並列要素の推定を行った．その結果を**表 9** に示す．表 9 では，正しい並列構造で単一並列構造と複合並列構造とに分類して正解数，不正解数，推定精度を記している．ただし，複合並列構造の評価では，親子関係，兄弟関係に含まれるすべての並列のキーに対して正しい並列要素を推定できた場合だけを正解とした．単一並列構造の推定精度は 76.4％（正解：331 個，不正解：102 個）であり，複合並列構造の推定精度は 72.2％（正解：394 個，不正解：152 個）であった．単一並列構造，複合並列構造のいずれの場合でも，正しい並列構造の前置要素と後置要素のいずれも文節数が 1 である並列構造の推定精度が高いことが分かる．並列のキーの抽出において第二種の誤りを犯すので，本手法による推定精度を考える場合，第二種の誤りも考慮しなければならない．そのため，全体を総合した推定精度は 70.5％（正解：725 個，不正解：254 個，第二種の誤り：49 個）となる．

並列要素の推定誤りの原因を，単一並列構造と複合並列構造のそれぞれについて調査した．その結果を**表 10** に示す．推定誤りの原因を出現頻度順に示すと以下のとおりになる．

原因 A： 4.3.2 項の［経験則 a］によって生じる誤り．

表 9 並列要素の推定結果
Table 9 Estimation result of the coordinate structure with our method.

	単一並列構造			複合並列構造			総計
正しい並列要素の文節数	1	2 以上	計	1	2 以上	計	
正しく推定	254	77	331	283	111	394	725
誤って推定	35	67	102	53	99	152	254
第二種の誤り							49
推定精度（％）	87.9	53.5	76.4	84.2	52.9	72.2	70.5

たとえば，「A の B と C」という文であれば，B を前置要素，C を後置要素とする．そのため，「(A の B) と (C)」が並列している場合は，推定誤りを犯す．

原因 B： 構造的類似性に基づいて並列要素を推定することで生じる誤り．たとえば，「A の B の C と D の E」という文であれば，(B の C) を前置要素，(D の E) を後置要素とする．そのため，「(A の B の C) と (D の E)」が並列している場合は，推定誤りを犯す．

原因 C： 並列要素を決定する表層的な手がかりではあるが，実際には並列要素を示していない場合に生じる誤り（**図 8** の誤り例 1）．この例文では，意味的には「プログラミング技術」と「計算機の能力」が並列である．しかし，後置表現の「など」が文末側にあるために，後置要素を誤って推定している．

原因 D，E： 前処理の解析誤りにより生じる誤り．文節の性質を誤って決定してしまうと，係り受け関係の決定にも誤りを生じてしまう．また，係り受け関係の解析誤りは，並列構造の最長範囲の決定に影響を及ぼす．

原因 F： 部分的並列を名詞並列と推定したことにより生じる誤り．部分的並列と名詞並列の存在を表す並列のキーに同じものがあるために起こる．

原因 G： 並列構造が重なり合う可能性がある場合は，4.4 節で述べたように推定順序を決定し，1 つずつ並列要素を推定する．したがって，推定順序が異なると，異なる推定結果が得られることがある．推定順序は並列のキーの特徴に基づいて決定したが，当然例外もある（図 8 の誤り例 2）．この例文では，「または,」の特殊キーが，その前後にある「,」よりも後に推定処理を行ったために推定を誤っている．

原因 H： 並列構造が重なり合う可能性がある場合は，1 つずつ並列要素を推定していく．したがって，並列要素の推定結果に誤りが生じると，後の並列要素の推定に影響を及ぼしてしまう．

これらの誤りのうち，原因 A，原因 B，原因 C のために起こる誤りがすべての誤りの約半数を占める（単一並列構造：64.7％，複合並列構造：43.4％）．並列要素の決定に際して［経験則 a］を用いたものが 324 個存在した．そのうち［経験則 a］を用いることで並列要素を正しく決定できたものは 275 個（正解率：84.9％）であった．また，構造の完全一致または部分一致によって並列要素を決定したものが 170 個あり，これにより正しく決定できたものは 125 個（正解率：73.5％）あった．さらに，並列要素を決定する表層的な手がかりが存在したものが 571 個あり，それによって前置要素または後置要素を正しく決定できたものは 533 個（正解率：93.3％）とかなり多かった．これらのことから，上記 3 つのものは並列要素の推定に際して有効であると考えられる．

表 10 推定誤りの原因
Table 10 Causes of estimation failure.

原因	単一	複合	計
A：[経験則 a] による誤り	26	23	49
B：構造的類似性による誤り	22	23	45
C：表層的な手がかりによる誤り	18	20	38
D：前処理の誤り	15	9	24
E：並列構造の最長範囲の推定誤り	6	4	10
F：並列構造が部分並列であった	1	1	2
G：並列要素の推定順序の誤り		15	15
H：別の並列要素の推定誤りの影響		36	36
その他	14	21	35
合計	102	152	254

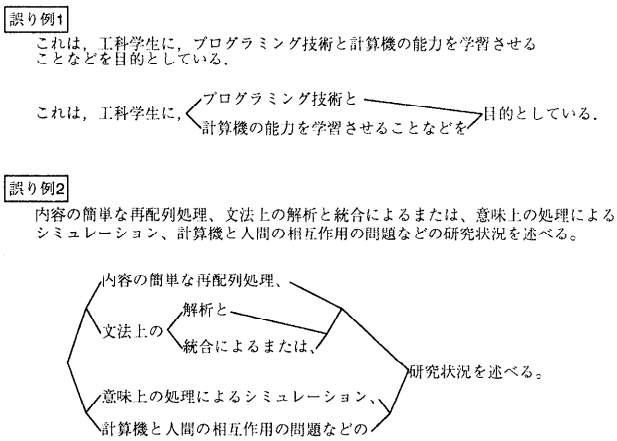


図 8 並列要素の推定誤りの例
Fig. 8 Examples of estimation failure.

本手法では，処理の高速化を計るために字面の情報だけを用いて文節切りや係り先の決定を行っている．表10に示した原因D，Eによる誤りは，この手法をとったために起こると考えられる．しかし，原因D，Eによる誤りは，今回の評価実験では34個（13.4％）しか現れず，誤り全体に占める割合は少ないといえる．

ユーザに煩わしさを感じさせない待ち時間で処理したいという要求から本手法のアプローチをとった．この要求を満たすことを確かめるためにSPARC station ELC（CPU：SPARC/33 MHz）で処理時間を測定した．その結果，上記の抄録文（54,858文字）から抽出するすべての並列構造を推定するのに約7.1秒を要した．また，別の約10万字の文章でも測定したところ，約14秒であった．このことから，上記要求を満たしていると考える．

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本研究の目的である．そのため，本手法は，単語の意味が分からない場合に読み手がとると思われる，並列要素の推定手順を模倣している．書き手の意図に合った推定結果が出力されなくても，読み手が読みとる可能性がある並列要素を推定し指摘できれば，その手法は推敲支援において有効である．その点から考えて，今回の評価実験で得られた推定精度（70.5％）は，本稿の処理の高速性から見て役立つ範囲内にあると評価している．

6. 推敲支援への適用

並列構造を含む文は一般に長くなり，並列構造の範囲やその前後の係り受け関係に曖昧性を含むことが多い．実際に，読み手が並列構造を含む文を読むとき，並列構造の曖昧性に直面することがよくある．このように，並列構造を含む文は，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．本章では，前章までに構築した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．

理想的な推敲支援とは，本当に問題がある箇所だけを指摘することである．しかし，並列構造が一意に定まると書き手が思う表現でも，読み手にその意図が伝わらない場合もある．本稿で述べた手法を用いて，文章中に含まれる並列構造の約97％を抽出することができた．そのため，文章中の並列のキーを抽出して指摘し，並列構造の再確認を書き手に促すだけでも，推敲支援として役立つと考える．また，文章を書いている場合，書き手は思い込んで書いている場合が多いので，違った角度から文章を眺める機会を提示することは有用である．そこで，並列のキーを指摘すると同時に，本手法を用いて得られる情報も書き手の求めに応じて指摘する．

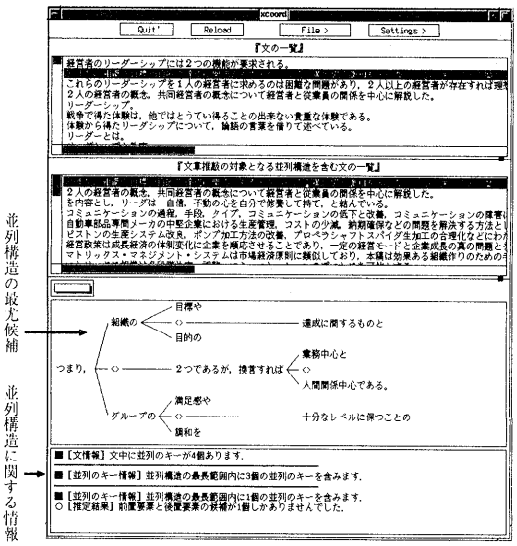


図9 推敲支援への応用
Fig. 9 Application to a writing tool for Japanese documents.

本手法の推敲支援への応用として，並列構造を指摘するシステムのプロトタイプをXウィンドウシステム上で作成した．**図9**にプロトタイプの表示例を示す．図中の上から2つの欄は，文の一覧，並列構造を含む文の一覧である．並列構造を含む文の一覧から1つ選択すると，下の2つの欄に解析結果を瞬時に表示する．ツリー状の図は推定した並列構造を表している．一番下の欄には，「並列要素を決定する表層的な手がかりが存在しない」や「並列要素の最長範囲が長い」など，並列構造を書き手が再考するときに参考になる情報を表示する．

5章で述べた調査結果をもとに1万字☆あたりの指摘数を換算すると178個となる．指摘数が多くなると，指摘されたものを1つずつ検討するのは煩わしくなる．そこで，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方法も用意している．この提示法とすべての並列構造を提示する方法とを書き手が選択できるようにする．

☆ 実用的な文章の大きさの目安としている．ちなみに本稿の本文は約14,000字である．

7. おわりに

本稿では，大規模な辞書を使わず，単語の意味を反映した解析や形態素解析を行わずに，名詞並列と部分的並列の並列構造を推定する方法について述べた．本手法では，並列要素を推定するために，並列要素を決定する表層的な手がかり，並列要素を決定する経験則，構造的類似性を利用した．構築した並列要素を推定する手法を文章に適用した結果，上の3つのいずれの決定規則も有効であることが分かった．さらに，並列構造の推定法を文章の推敲支援へ応用する方法に関して述べた．

今後の課題としては，まず，並列要素を決定する規則の充実があげられる．文章の推敲支援に応用する場合に，複雑な並列構造を誤って推定してしまっても，読み手に誤って受け取られる可能性があることを書き手に指摘できれば有用である．しかし，明らかに解析誤りであると分かるものを指摘すると書き手は煩わしく感じる．また，本稿では，名詞並列と部分的並列を対象とした並列構造の推定について述べた．この方法を述語並列の推定へ拡張することも今後の課題である．さらに，推敲支援への応用としてプロトタイプを作成した．このプロトタイプを改良して，推敲支援ツールとして完成度をあげることも今後の課題である．

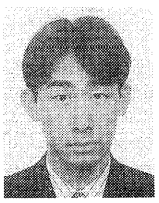
参考文献

1) 田村直良，田中穂積：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報告，No.59, pp.1–8 (1987).
2) 黒橋禎夫，長尾　眞：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) 首藤公昭，吉村賢治，津田健蔵：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).
4) 吉田　将：二文節間の係り受けを基礎とした日本語文の構文分析，電子通信学会論文誌，Vol.55-D, No.4, pp.238–244 (1972).
5) 首藤公昭，楢原斗志子，吉田　将：日本語の機械処理のための文節構造モデル，電子通信学会論文誌，Vol.62-D, No.12, pp.872–879 (1979).
6) 山上晃司，安原　宏：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，No.98, pp.9–16 (1993).
7) 菅沼　明，笠原　晋，牛島和夫：字面解析による用言の活用形推定について，情報処理学会論文誌，Vol.37, No.6, pp.1007–1016 (1996).
8) 黒橋禎夫，長尾　眞：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，No.88, pp.1–8 (1992).

(平成8年10月9日受付)
(平成9年 5 月8日採録)

菅沼　明（正会員）

1961年生．1986年九州大学工学部情報工学科卒業．1988年同大学大学院工学研究科情報工学専攻修士課程修了．1991年同博士後期課程修了．同年九州大学工学部情報工学科助手勤務．1993年同大学工学部情報工学科講師，1996年同大学大学院システム情報科学研究科助教授，現在に至る．工学博士．日本語情報処理，ユーザインタフェース，遠隔講義支援環境，ニューラルネットワークの応用などに興味を持つ．1994年情報処理学会奨励賞受賞．日本ソフトウェア科学会会員．



山村　広臣（正会員）

1972年生．1994年九州大学工学部情報工学科卒業．1996年同大学大学院工学研究科情報工学専攻修士課程修了．同年NTTデータ通信（株）に入社．日本語情報処理，ネットワーク通信技術に興味を持つ．

牛島　和夫（正会員）

1937年生．1961年東京大学工学部応用物理学科（数理工学コース）卒業．1963年同大学大学院修士課程修了．同年九州大学中央計数施設勤務．1977年九州大学工学部情報工学科教授（計算機ソフトウェア講座担当），1996年同大学大学院システム情報科学研究科教授，現在に至る．1990年4月から1994年3月まで九州大学大型計算機センター長，1996年から九州大学大学院システム情報科学研究科長を兼務．1991年度情報処理学会九州支部長．1995/96年度本学会監事．工学博士．電子情報通信学会，ソフトウェア科学会，ACM，IEEE-CS各会員．