従属節の依存関係を考慮した日本語係り受け解析について

白井　諭[†]　横尾昭男[†]　木村淳子[‡]　小見佳恵[‡]

[†]ＮＴＴコミュニケーション科学研究所　[‡]ＮＴＴアドバンステクノロジ

1 はじめに

自然言語の構文解析の方法として，従来から多くの方法が提案されてきた。日本語に対しては，語順入れ替わりや省略などに強い係り受け解析の方法が適していると考えられる[長尾93]が，長文に対しては必ずしも良い成績は得られていない。

最近，表層的な特徴を利用して長文にアプローチするなどの試みが始められ[黒橋92, 黒橋94, 奥村93]，並列関係の解析などには有効であることがわかってきた。また，呼応関係などの文構造を決定する要因に着目して係り受けの曖昧さを局所化する方法も提案されている[池野93]。このように，従来の日本語の構文解析では，日本語の表層的特徴を十分には使いきっていないと考えられる[長尾93]。

長文に対する係り受け解析が失敗する主な要因としては従属節の相互関係と名詞句の並列関係の解析失敗が挙げられる。そこで筆者らは，南による従属句の分類[南74][南91]を改良・詳細化した従属節述語の相互関係の解析方式と，単語の意味属性や構文的特徴に基づく部分並列の認定も可能な名詞句の並列関係の解析方式を提案し[白井93, 白井94a]，処理系の試作によりその有効性を確認した[白井94b]。

本稿では，本格的な処理系実現の第１ステップとして，試作した処理系を見直し，４つのルールグループとその駆動系および若干の個別処理に再構成するとともに，ルールの必要量について検討した結果について報告する。

2 係り受け解析方式の概要

筆者らが提案した係り受け解析は次の５つのステップにより構成される。

①文節の連接を見て述語句の認定と分類を行なう
②分類に基づいて述語句間の係り受けを決定する
③形式と類似性などに基づき並列関係を検出する
④部分的な並列関係があれば文節を適宜分割する
⑤多項関係により連体・格・副詞修飾等を決定する

2.1 述語句の認定と分類

日本語研究の立場から，南は構文的特徴や意味的関係に基づいて従属節をＡＢＣの３種類に分類することを提案し，Ａは従属性が強くＣは独立性が強いといった構文上の強い傾向があることを指摘している[南74][南91][南93]。しかし，自然言語処理において意味的関係を構文解析の段階であらかじめ考慮するのは一般に困難であるため，これまでは日本語の構文解析へはうまく適用されていなかった。

これに対して筆者らは，表層的な情報のみによる従属節述語の分類について検討し，表１のような52分類(13×4)を提案した[白井93][白井94a]。これは，

表１　述語句の分類

形態的分類		機能的細分類		備　考
従属節	連用節	A		～ｼつつ，～ｼながら(=継続) ～ｽﾙのに｜続いて ～ｽﾙことに｜加えて
		A＋読点		
		B	通常	＜通常＞ ～ｼ，～ｼて，～ｽﾙので ～ｽﾙため，名詞＋で(動詞)
			強中止	
		B＋読点	通常	＜強中止＞ ～ｼており，～ｽﾙことで ～ｽﾙものであり
			強中止	
		C		～ｽﾙが，～ｽﾙし
		C＋読点		
	引用節	引用相当		～ｽﾙよう(依頼する) → B
		引　用		～ｽﾙと(発話する) → C＋読点
	連体節	限定修飾		一般名詞へ係る → B
		捉え直し		形式名詞へ係る → B＋読点
主　節		―――		文末の述語句
動作性		名詞性述語		名詞＋指定の助動詞
各従属節や主節の機能的細分類に適用する		形容語性述語		形容詞，いわゆる形容動詞
		自動詞性述語		自動詞，受身を伴う他動詞
		他動詞性述語		他動詞，使役を伴う自動詞

A Japanese dependency analysis that considers the relations between subordinate clauses

Satoshi SHIRAI[†], Akio YOKOO[†], Junko KIMURA[‡] and Yoshie OMI[‡]

[†] NTT Communication Science Laboratories and [‡]NTT Advanced Technology Corporation

同時や継続の表現をA，条件や原因理由や中止などをB，独立的なものをCにそれぞれ割り振った後，述語の性質に応じて細分類したもので，このうちABCの3分類については，南の分類のうち表層情報で分類可能なものは尊重されているが，分類困難なものは基本的にはBに分類される結果となっている。ただし，逆接の「～するが」は意味的にはBに分類すべきであろうが，新聞記事等では順接にも逆接にも解釈可能なものが多用されることから，簡単のためCに分類することにした。また，「～するのに続いて」「～することに加えて」など複数の文節により構成される表現を「述語句」としてAに分類するなど意味的なまとまりを扱うようにし，南が指摘した構文的特徴の保存を図った。

述語句の分類はルールとのパタン照合により行なうこととし，新聞記事300文に対しては54ルールが抽出されている。

2．2 述語句の係り受けの決定

前節の述語句の細分類を段階的に適用することにより述語句間の係り受けを決定する[白井94a]。まず，ABCおよび読点を考慮した6分類により係る／係らないを決定する。即ち，A（読点なし）が最も依存性が強く，以下，A＋読点，B（読点なし）の順で続く。逆に，C＋読点が最も独立性が強く，以下，C（読点なし），B＋読点の順で続く。このとき，同じ分類に属する述語句同士は係る／係らないが不明であるため，その下位分類の適用により決定を試みる。例えば，B＋読点同士に対しては強中止による決定を試み，それでも決定できなければ動作性による決定を試みる（名詞性述語の依存性が最も強く，逆に，他動詞性述語の独立性が最も強い）。動作性を考慮しても係る／係らないが決定できない場合は，係る場合と係らない場合を多義として展開することにより対処する。

追跡調査の結果，上記の方法により，文単位で見たときの述語句間の係り受けの正解率はほぼ100%となることがわかった。なお，動作性を考慮しても係る／係らないが決定できない場合に，多義展開を行なわず直近への係りのみを用いるようにしても述語句間の係り受けの正解率は98.4%が見込まれる。

述語句の係り受けのルールは新聞記事300文に対しては34ルールが抽出されている。

2．3 並列関係の検出

並列関係にある名詞句は，それが使用される文脈によって規定される観点において共通性を持つと考えられる。従って，解析システムで用いられている単語属性だけでは十分な解析が行えないと予想されるが，並列として用いられる観点をあらかじめ網羅的に用意しておくのは不可能である。しかし，新聞記事等の調査から，実際には特殊な観点が使われることは少なく，表層情報や解析システムで設定されている単語属性（意味属性）並びに構造的な特徴を組み合わせれば並列関係がほぼ正しく検出できるという見通しを得たので，筆者らは次のような並列の解析手順を提案した[白井93]。

①構成要素として同じ表記の単語が使われている
②構成要素として意味属性の類似した単語がある
③並列マーカや品詞構成など構造的な特徴を持つ
④以上に該当しなければ直近の文節の並列とする

上記により解析可能な並列の例と件数を表2に示す。ただし，部分並列については次節で検討する。

表2　並列の認定基準（新聞記事500文の該当数）

認定基準	件数	新聞記事の例
表記の同じ単語	13	地上七階、地下一階
類似の意味属性	39	米国、東南アジア
構造的特徴あり	33	ビデオテープ、テキストなど
直近文節の並列	18	海外の現地法人や～社の販売ルート
部分並列を含む	10	日本、米国市場
現状では困難	1	近視、遠視、弱視の視力回復と視力低下を予防するプログラム

今回は，上記の手順をルールにより記述するという方針で，試作した処理系の再構成を実施した。ただし，名詞止めのような述語省略に伴う並列構文は形式上は名詞句の並列と同形であり，切り分けるためのルールの記述が難しい場合がある。このため，

タイプ別にいくつかの個別処理を作成した。

並列を検出するためのルールは，次節の部分並列処理と併せると，新聞記事300文に対し30ルールが抽出されている。

2．4 部分並列の処理

部分並列とは，形態素解析により出力される文節を単位としてみたとき，例えば「日本、米国市場」や「ＪＩＳ第一、第二水準」のように，文節全体同士では並列関係にないものをいう。実用文にもしばしば現れるが解析は困難であるとされてきた。新聞記事等の調査によれば，このタイプの並列も単語表記の一致や意味属性の類似など，表層的な特徴で検出可能なものが９割を占めることがわかった。部分的な並列を検出した場合，図１のように検出した並列の前後で文節を切断し，連体関係など適当な係り受けリンクを張ってやれば，以降は文節ごとの並列と同様の処理が可能になる。

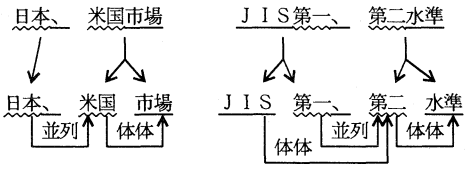


図１　部分並列の解消

2．5 多項関係による係り受けの決定

多項関係が有効となるのは，例えば，「どんな/戦略を/選択するか」において「戦略を」を介した「どんな」と「か」の呼応関係を捉えたり，「社員の/半数は/～」において「半数」に「社員」の持つ意味属性を継承させたりする場合である。ただし，大多数の係り受けに関しては２項関係のルールで十分記述可能である。

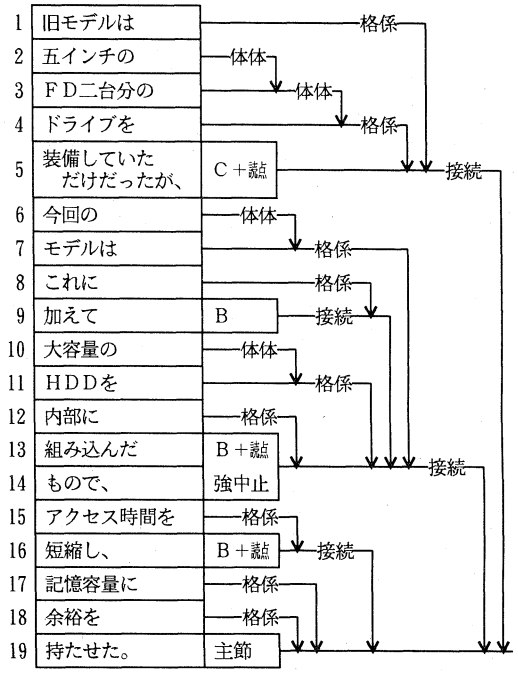
新聞記事300文に対しては158ルールが抽出されている。

3 提案方式の到達点

提案方式による係り受け解析の例を図２に示す。本方式により，従属節の多い文や部分並列を含む文など，これまでは解析困難とされてきた文も解析できるようになった。

新聞記事300文（日経産業新聞，情報欄リード文）

＜例文＞　旧モデルは五インチのＦＤ二台分のドライブを装備していただけだったが、今回のモデルはこれに加えて大容量のＨＤＤを内部に組み込んだもので、アクセス時間を短縮し、記憶容量に余裕を持たせた。（９１字，従属節の多い文）



＜例文＞　情報事業、海外技術両本部内に海外企画部という統括部をつくり、自社ブランド製品の輸出拡大やソフトウエアの海外現地生産などを集中管理する。（６７字，部分並列を含む文）

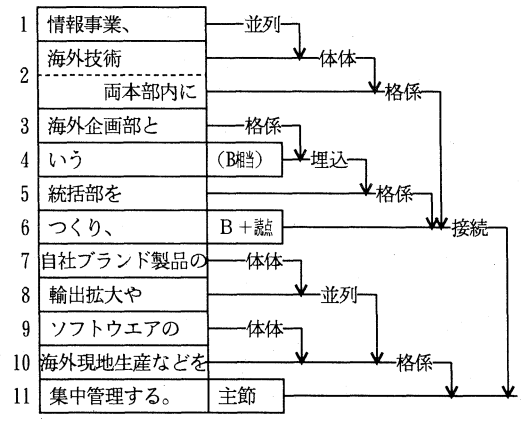


図２　係り受け解析の例

表3 提案方式の解析精度

文字数	文数	1位正解	1位同点	正解含む	多義
6～10	9	9(100%)	0	9(100%)	1.1
11～20	17	17(100%)	0	17(100%)	1.1
21～30	40	39 (98%)	1 (2%)	40(100%)	1.9
31～40	51	48 (94%)	3 (6%)	51(100%)	3.2
41～50	58	52 (90%)	4 (7%)	58(100%)	5.2
51～60	48	41 (85%)	5 (10%)	48(100%)	7.8
61～70	42	36 (86%)	2 (5%)	42(100%)	10.3
71～80	19	16 (84%)	2 (11%)	19(100%)	11.7
81～90	10	9 (90%)	1 (10%)	10(100%)	19.0
91～100	3	3(100%)	0	3(100%)	28.3
101～111	3	3(100%)	0	3(100%)	16.0
合 計	300	273(91%)	18 (6%)	300(100%)	10.3

に対する走行結果を表3に示す。表から，文が長くなっても高い割合で1位正解が得られ，300文全体では91%の1位正解率となっている。また，1位正解が得られなかった文も，評価点で見れば1位と同点のものが多く，残りも数位以内に正解が含まれており，本方式が十分高い精度を持つことがわかる。ただし，文が長くなると，特に並列解析に伴う多義の発生が多く，この多義の削減が今後の課題である。

新聞記事300文の解析に要したルール数をルールグループごとに集計し，文数との関係を併せて調査した。結果を図3に示す。図からルール数は飽和傾向にあり，少なくとも新聞記事を対象とする限り，数百～1000文で飽和すると予想される。

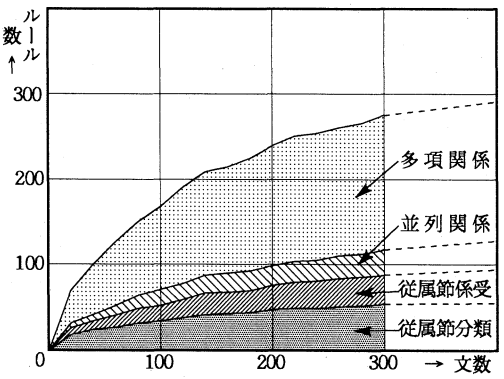


図3 文数とルール数の関係

4 おわりに

本稿では，先に提案した従属節述語の相互関係と，名詞句の並列関係の解析機能を強化した係り受け解析方式に対し，処理系を見直すとともに，現状の到達点について検討した。具体的には，試作した処理系を4つのルールグループとその駆動系および若干の並列構造を解析するための個別処理に編成し，新聞記事300文に対する走行実験を行なった結果，1位正解率が91%であること，多義を許容すれば数位以内に正解が含まれること，また，数百～1000文を対象とすればルール数は飽和する見込みであることなどを報告した。

今後は，対象文数を増やしてルールの整備を進めることにより係り受け解析処理の完成を目指す。また，現状の問題である並列解析に伴う多義の削減についても検討する予定である。

＜謝辞＞
係り受け解析処理の実現にご協力くださった松尾三津恵氏，中村三紀氏を始めとするＮＴＴアドバンステクノロジの各位に感謝する。

＜参考文献＞

[池野93] 池野，奥村，松下，山本，永田：日本語長文の翻訳における副詞呼応範囲の優先構造化方式，第7回人工知能学会全大17-6
[黒橋92] 黒橋，長尾：長い日本語文における並列構造の推定，情処論Vol.33 No.8
[黒橋94] 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，自然言語処理Vol.1 No.1
[南74] 南：現代日本語の構造，大修館書店
[南91] 南：現代日本語の従属句についての小調査，明治書院，日本語学Vol.10 No.12
[南93] 南：現代日本語文法の輪郭，大修館
[長尾93] Nagao, M. : Varieties of Heuristics in Sentence Parsing, Invited Talk at International Workshop on Parsing Technology, Tilburg/Durbuy
[奥村93] 奥村，池野，松下，山本，永田：日本語文の並列構造を利用した長文解析方式，第7回人工知能学会全大17-4
[白井93] 白井，横尾，木村，小見：日本語従属節の依存構造に着目した係り受け解析，第47回情処全大3M-1
[白井94a] 白井，横尾，木村，小見：日本語従属節の相互関係に関する一考察，1994年春期信学全大D-116
[白井94b] 白井，横尾，木村，小見：従属節の依存関係を考慮した日本語係り受け解析の精度，第49回情処全大1G-10