

## 述語の隣接と述部の付加における意味的階層性

大谷 朗\*† 宮田 高志\* 松本 裕治\*

奈良先端科学技術大学院大学情報科学研究科\* 大阪学院大学情報学部†

{akira-o, takashi, matsu}@is.aist-nara.ac.jp\* ohtani@utc.osaka-gu.ac.jp†

HPSGのような制約にもとづいた文法は、自然言語処理において効果的である。汎用的な文法体系を構築するため、我々は南の統語的/意味的階層という考えを部分的に実装した、NAIST JPSGという三層句構造文法を構築した。NAIST JPSGでは、日本語の様々な特徴の検討を経て、それらの間の規則性を局所的制約として記述することで、文法の原理、スキーマおよび素性が設計されている。三つの層は、日本語の文が、その統語情報を描叙し (Layer I)、意味情報を指定し (Layer II)、そして文脈情報を文の中核に付加する (Layer III) ことで形成されるさまを反映している。また、我々は「ながら」と「ので」に導かれた従属節の分布について、素性の集合として記述されるそれらの語彙情報に関心を持って分析した。

[キーワード] HPSG, 汎用的な文法体系, 三層句構造文法, NAIST JPSG, 南の統語的/意味的階層, 従属節

## Semantic Hierarchy for Adjacency of Predicates and Adjunction of Predicative Constituents

Akira Ohtani\*† Takashi Miyata\* Yuji Matsumoto\*

Graduate School of Information Science, Nara Institute of Science and Technology\*

Faculty of Informatics, Osaka Gakuin University†

A constraint-based grammar like HPSG is effective in natural language processing. For constructing a general-purpose grammar system, we develop three-layered phrase structure grammar, NAIST JPSG, which is a partial implementation of ideas from Minami's four-level syntactic/semantic hierarchy. The principles, schemata and features are designed through considering various aspects of Japanese and describing regularities among them as a set of local constraints. The three layers reflect the manner the Japanese sentence is formed by describing its syntactic part (Layer I), specifying its semantic part (Layer II) and adding some contextual part to its core (Layer III). We then analyze the distribution of subordinate clause headed by *nagara* and *node*, with main focus on their specific lexical information, which is described as a set of features.

[Keyword] HPSG, general-purpose grammar system, three-layered phrase structure grammar, NAIST JPSG, Minami's syntactic/semantic hierarchy, subordinate clause

### 1 はじめに

従来の言語処理に関連した日本語文法は、それを利用するシステムが当面のデータに対する被覆率や解析効率の向上を指向していたため、文法自体は「アプリケーションより」に特化されることが多かった。そのような方針がとられた背景には、単に

利便の追及という点からの要請だけでなく、対象とする言語データの性質が解析に利用できるほど明確にされていない、という事情もあった。

統語解析の役割が文の構造を特定の意味表示に写像するだけなら、(形態素解析により) 単語や句・節の境界さえ正しく解析できれば、どのような統語構造(を構築する理論)を考えたところでさしたる

効用は得られない。実際、多くのシステムはこの方向性の追及をあきらめ、解析途中に不足する情報を別の処理から補うことで文法を非常に単純化してしまっている。しかしながら、そのような考え方は各処理間の整合性を保つことが難しく、データの変更や規模の拡大などへの対応は決して柔軟なものとは言えない。また、このような問題は昨今のものではなく、統語と意味の融合処理が望まれつつも実現されないまま今日に至っている。

文法の特化という考え方は、限定されたデータを処理する上では有効なアプローチであることは否定し得ない。しかし反面、広汎なデータ、すなわち言語を持つ全体的な性質を見失いがちであり、将来必要とされる、汎用的な処理には欠くことのできない「見通しのよい文法」の形式化の研究からは乖離してしまうことも想像に難くない。

そこで、本論文は、日本語の階層的な表現過程について提案された南の階層構造 [12, 13] を導入し、文解析の様々な話題について句構造文法による実装可能な形式化を試みることにする。日本語の全体的な性質を巨視的にとらえた南の階層観は、そのままでは処理に適用することはできないが、意味的性質を軸に述語の隣接を制約すると、文構造だけでなく述部の付加といった係り関係までもが局所的關係として規定されていることがわかる。そのような局所性に依拠し、本論文ではさらに言語処理に有効な統語・意味（文脈）情報が統合された汎用的な言語情報構造の構築を目指すことにする。

## 2 日本語の階層構造と句構造文法

### 2.1 日本語の階層構造

南 [12, 13] は、「書き手の対象に対する認識とその表現過程は文の階層的な構造に反映される」といった日本語文法に関する従来の研究を発展させたものである。南の提案する階層は、主に従属節における構成要素の共起関係にもとづいて、ACBC<sub>C</sub>CD という包含関係にある四つの分類を仮定している。各類の要素は同類の要素にも含まれ、その結果、文の構成要素は重層的な関係をもつことになる。(1) にそのような階層から説明できる現象の一例を示す。

(1) a.\*<sub>[A類 [B類 健が] ナイターを見ながら]</sub>  
奈緒美は ラジオを 聞いた

b. <sub>[B類 [B類 健が] ナイターを見ると]</sub>  
奈緒美は ラジオを 聞いた

- (3) B 類節 = ... 主格名詞句 A 類節 アスペクト...  
 (4) A 類節 = 様態修飾句\* 非主格名詞句 本動詞...  
 (9) 様態修飾句 = A 類従属節 | 様態副詞句 | 頻度副詞句  
 (10) A 類従属節 = A 類節 A 類節従属節ヘッド

図 1: 吉本 [8] の A 類に関する句構造規則の抜粋

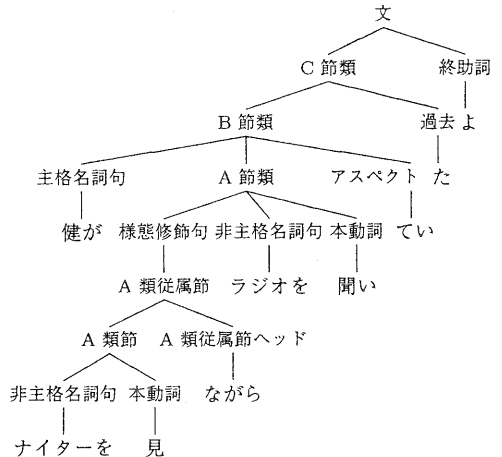


図 2: 図 1 にしたがった (3a) の句構造木

接続助詞「ながら」と「と」に導かれる従属節は、それぞれ A、B 類と分類されるが、B 類の要素である主格名詞句「健が」は先の制限により A 類には含まれないため、(1) のような文法性が生じる。<sup>1</sup>

また、主語の一つがゼロ化した (2) も、その構造が (3a) のような主格名詞句を主節の要素としたものであることを、この階層は説明することができる。

(2) 健がナイターを見ながらラジオを聞いていたよ

(3) a. <sub>[主節 健が [従属節 ナイターを見ながら]</sub>  
ラジオを 聞いていたよ]

b.\*<sub>[主節 [従属節 健が ナイターを見ながら]</sub>  
ラジオを 聞いていたよ]

吉本 [8] は、南の階層観の一部を図 1 に示すような句構造規則として形式化した。吉本の句構造規則は、各類の節を導く従属節がそれ自体、それと同じ類の中にその構成素として埋め込まれるという点で、再帰的な構造をとっている。図 2 にそのような分析による (3a) の句構造を示す。

<sup>1</sup> 「ながら」には順接 (A 類) と逆接 (B 類) の二つの用法があり、逆接の解釈が可能であれば従属節中には主格名詞句 (B 類) は含まれる。[12, 2]

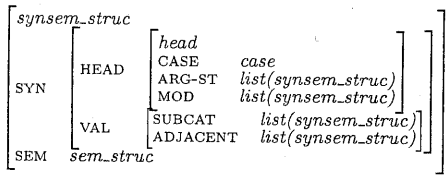


図 3: NAIST JPSG の素性構造

$$\text{日本語} = P_1 \wedge \dots \wedge P_n \wedge P_{n+1} \wedge \dots \wedge P_{n+m} \wedge (S_1 \vee \dots \vee S_p \vee L_1 \vee \dots \vee L_q)$$

図 4: HPSG にもとづく日本語文法の形式化

吉本は後に、主辞駆動句構造文法 (Head-driven Phrase Structure Grammar: HPSG) [4] にもとづくことで、図 1 のような句構造規則を廃している。[9] しかしながら、そのモデルにおいても、南に做った四分類の階層関係のみが形式化され、「対象に対する認識とその表現過程」といった意味的側面は考慮されていない。<sup>2</sup>

そこで、本論文は吉本らの形式化 [9] を補う形で、より充実した内容をもった階層構造のモデルを考えるのであるが、その具体的記述に先がけて、2.2 節では基盤となる枠組について述べることにする。

## 2.2 日本語句構造文法

NAIST 日本語句構造文法 (Japanese Phrase Structure Grammar: JPSG)<sup>3</sup> は、当面の記述対象を日本語に限定しているものの、HPSG[4, 6] に準じた、いわば「言語理論より」の文法である。[3]

NAIST JPSG は、言語に普遍的な性質と日本語の性質を切り分けることで、HPSG の普遍原理 ( $P_1, \dots, P_n$ )、個別言語の原理 ( $P_{n+1}, \dots, P_{n+m}$ )、選言的スキーマ ( $S_1, \dots, S_p$ )、および概ね図 3 のような素性構造にもとづいた語彙情報 ( $L_1, \dots, L_q$ ) の単一化によって、日本語に特徴的な言語事実 (の集合、すなわち日本語) を図 4 のように形式化する。

詳細は別稿 [3] にゆだねるが、例えば使役文「健が奈緒美に本を読ませた」の (部分的な) 素性構造は図 5 のように記述することができる。

NAIST JPSG は、無効な構文木の生成や曖昧性の抑制などといった実装における諸問題を考慮に入れて、HPSG に対し独自の拡張を施した理論的枠組である。しかし、基本的には局所的に記述された

<sup>2</sup> ただし、形式化の主たる関心は日本語複合文における時制の説明にあり、階層に依拠しない形でモデルを發展させている。

<sup>3</sup> ICOT による JPSG[1] と区別してこのようによぶ。

(HPSG に準じた) 諸制約と、原理からは導き得ない語彙情報を蓄積した辞書、そしてそれらの間の単一化といった一般的な計算機構により、文法現象の体系的な説明を目指すものである。このような考え方によれば、「特殊な」現象の処理を指向した「特殊な」計算機構の導入はもはや必要なく、また設計が単純化されることで、システムの見通しが良くなることを期待したモデルでもある。

## 2.3 南の階層構造の形式化に向けて

A, B, C, D といった恣意的なラベルに依ってはいるが、吉本らの研究は階層構造の統語的側面の局所性を明確にしている。南の巨視的な階層観は、そのままの記述では有効な実装が困難であるが、本論文では「統語・意味情報の統合構造モデル」として、NAIST JPSG の観点から、意味的側面についても以下の方針で形式化を試みることにする。

- I. 文構造が表示する統語・意味情報は、それを構成する語彙に記載された情報のみである。
- II. 語彙を構成素とした構造、そのような構造を構成素としたより大きな構造は、構成素の統語・意味情報が互いを制約することで構築される。
- III. 制約は局所的に規定される。

つまり、局所的に制約された統語情報にもとづく文構造の構築では、意味情報がさらに制約を課し、また意味情報を主導として構築がなされるモデルを提案するのである。また、このような方針によれば、図 3 の語彙記述における *sem\_struct* (意味素性構造) の拡充を中心に、その制約を図 4 に追加することで NAIST JPSG を拡張することになる。

そこで、3 節では、*sem\_struct* を記述するために、日本語の階層構造でとらえられている意味的側面について考察することにする。

## 3 階層構造が意味する意味

### 3.1 階層観から階層構造へ

日本語は、文末に見られる豊かな表現や係り関係などの修飾構造の分析からも明らかのように、文の構造の一般的制約としては統語情報よりむしろ意味情報の寄与するところが大きい。

文の意味的側面が何らかの階層を反映しているといった「階層観」自体は特に新しいものではない。山田 [10] 以降、いくつかの研究にその先例を見ることができ、それらを包括する形でより多くの



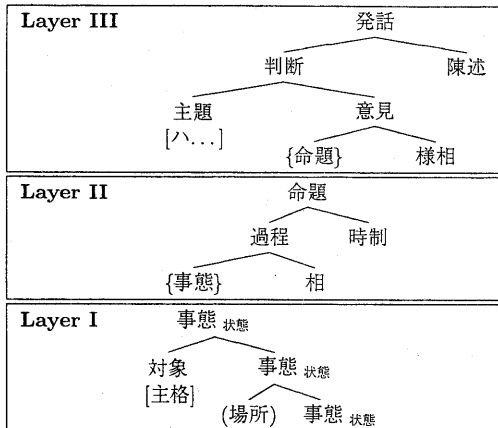


図 7: Layer にもとづく NAIST JPSG の階層モデル

#### 4 Layer にもとづく階層モデル

NAIST JPSG は階層構造を統語・意味（・談話）情報の‘インターフェース’として形式化している。各情報の回合点は語彙だけでなく、それを構成素としたより大きな構造にまでおよぶが、全ての情報は局所的に制約されている。3.2 節で述べたように、各 Layer はその主たる関心によって扱うべき意味的性質が異なっているが、単一化により統合された情報は Layer 間を伝播する。

以下、そのような Layer の特徴について述べる。

##### 4.1 Layer I: 述語的語彙情報（事態の描写）

Layer I は述語とその項構造により文の意味の中核を成すものである。sem\_struct が項構造素性 (ARG-ST) を介して下位範疇化素性 (SUBCAT) と連携し、述語が表わす事態とその参加者を制約する。<sup>4</sup>

各述語の項構造は、基本的には語彙情報の集積である辞書に、各々に特有な情報として記載される。しかしながら、項構造のタイプは述語の数の膨大さに比べると非常に限られている。

そこで、本論文では図 8 に示すような四分類を提案する。このような分類は、述語が表す相といった意味的側面からだけでなく、「～ている」や時間を表す副詞などとの共起関係といった統語現象の説明においても重要である。[11]

状態述語を例に Layer I における動詞の sem\_struct を図 9 に示す。<sup>5</sup> 相・時制に関する情報を指定する

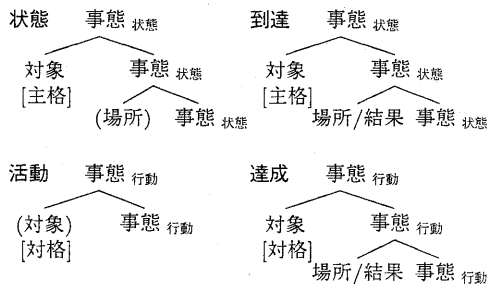


図 8: Layer I に現れる述語の階層情報の典型

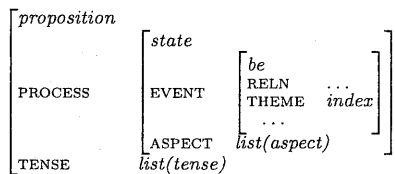


図 9: Layer I における状態述語の sem\_struct

素性 ASPECT および TENSE の値をリストとしているが、これは、動詞本来の情報が Layer II の要素を伴うことで、そのような情報についてのさらなる指定を受けることを記述したものである。語彙に固有な相（・時制）の情報をどう記述するかは現象の分析を含めた今後の課題であるが、図 8 に示した典型的な述語は図 10 の型階層を遵守している。<sup>6</sup>

また、使役や受動などの Layer I に現れる様態辞の主たる機能は、複合述語形成における項構造の変化に連携した階層構造の合成に他ならないが、例として使役述語の sem\_struct を図 11 に示しておく。

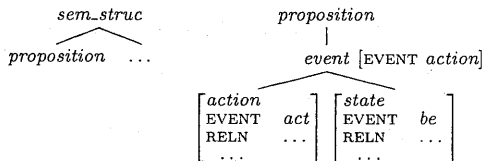


図 10: sem\_struct の構成に関する型階層

<sup>6</sup> 図 8 の分類に対応して、各述語の意味表示に以下のような概念構造的記述を導入することを考えている。

状態 [y BE<state>/AT<place>]  
 到達 [BECOME [y BE<state>/AT<place>]]  
 活動 [x ACT (on) y]  
 達成 [[x ACT(on) y] CASUSE [BECOME [y BE<state>/AT<place>]]]  
 概念的意味に関しては、現在も活発な議論がなされているが、NAIST JPSG では生成辞書 [5] の考え方を参考に、これらの分類に収まらない述語の意味記述を派生する方法を検討している。しかしながら、このような記述を導入するには議論の余地が残されており、本論文では図 9 のような素性記述にとどまった。

<sup>4</sup> 2.2 節を参照のこと。

<sup>5</sup> 名詞の sem\_struct(index) についても同様の記述が必要であるが、現時点では具体的に記述できていない。脚注 6 も参照。

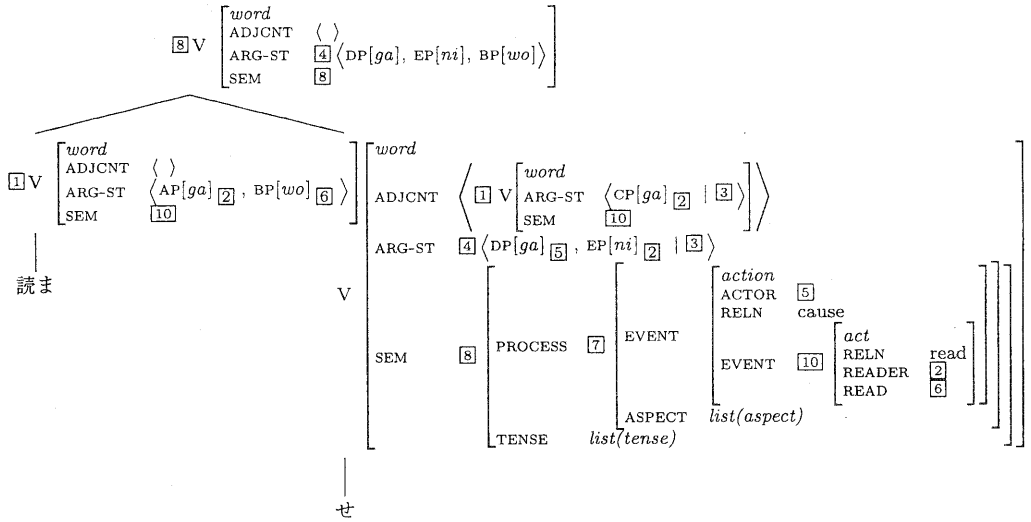


図 11: Layer I: 「読ませ(る)」の素性構造

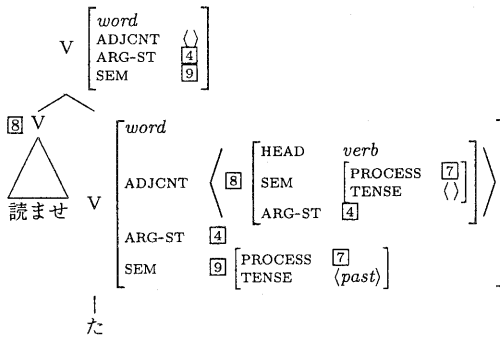


図 12: Layer II: 「読ませた」の素性構造

#### 4.2 Layer II: 指定的語彙情報 (命題の認定)

Layer II の機能は、Layer I が規定した内容をさまざまな意味の側面について認定することにある。4.1 節に関連すれば、その主な機能は相・時制情報を合成的に計算することであり、例えば図 11 の素性構造を、時制辞を含む Layer II のレベルで記述しようとする、概ね図 12 のような構造を仮定することになる。

ただし、語彙に固有な情報の記述の問題に加え、相(・時制)辞がさらに指定するような情報の合成は、リスト要素を単純に接続すればいいというわけではなく、やはり今後の課題となっている。<sup>7</sup>

<sup>7</sup> [11] など参照。なお、ICOT JPSG では「視野」という独自の素性を用いて、動詞に固有な情報の情報だけでなく、合成的な相素性についても明解な説明を与えている。NAIST JPSG

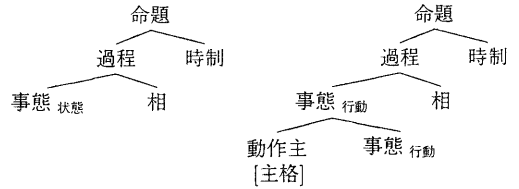


図 13: Layer II の階層情報の典型

時制辞に関しては、意思的な事態(行動)における主格名詞(動作主)が時制と共起関係にあるのに対し、非意思的な事態(状態)における主格名詞(対象)は時制の区別がない節にも現れる、ということが記述できなければならない。

そこで、本論文は図 13 に示す分類を提案する。また、意思的な事態を表わす活動述語を例に、Layer II における動詞の *sem\_struc* を図 14 に示す。図 14 は図 9 同様に図 10 に示した形階層を遵守している。

#### 4.3 Layer III: 付加的機能情報 (発話の提出)

Layer II までの情報は、Layer I (の述語が内包するデフォルト)の相・時制情報などを単一化によって指定することで計算されている。その主な関心は、文の統語的・意味的側面にあったが、Layer III では、さらに文脈的な情報が記述の対象となる。

では、そのような説明を図 8 に示した述語の階層情報や脚注 6 でふれた生成辞書的な考え方に関連付けようとしているが、現時点では記述できていない。

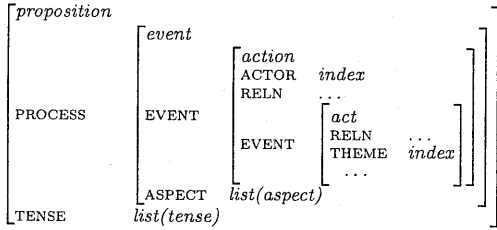


図 14: Layer II における活動述語の *sem\_struct*

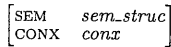


図 15: *sem\_struct* と区別された *conx* の位置付け

例えば、「は、なら、…」などの提題助詞を伴う主題要素は、従属節内には収まりきらず、文末での言い切りを要求することが知られている。

- (4) a. 中田が蹴ったボールを止めた
- b. 中田は蹴ったボールを止めた

(4a) には、(i) 中田がトラップした、(ii) キーパーがセーブした、という二つの解釈があるのに対し、(4b) は (i) の解釈しかない。2.1 節の議論と同様に、(4a,b) の‘階層’構造的な違いは (5) のように主語が補えるか否かで確かめることができる。

- (5) a. 中田が蹴ったボールをキーパーが止めた
- b. \*中田は蹴ったボールをキーパーが止めた

このように主題要素が文内に収まっている場合であれば、Layer 間の情報の伝播として記述することも可能であるが、「～は」が関係する範囲は、文脈の許す限りにおいては他の文にまで及ぶこともある。

また、さまざまな様相や陳述辞を伴うことで表される、Layer II で指定された情報 (命題) に対する話者の意見や判断といった発話内容は、どのように記述するのか、現在のところ明確になっていない。

よって、本研究では、図 15 に示すように、これらの文脈的性質を Layer I, II が規定する *sem\_struct* とは区別し、文脈素性 (CON(TE)X(T)) として記述することを提案するにとどめておく。

## 5 従属節の分布と意味的階層性

4 節で論じた階層モデルにもとづくと、統語構造と意味的階層性の対応関係を考えることで、様々な現象が局所的な制約として説明できる。5 節では、接続助詞「ながら」と「ので」を例に、そのような現象について述べることにする。

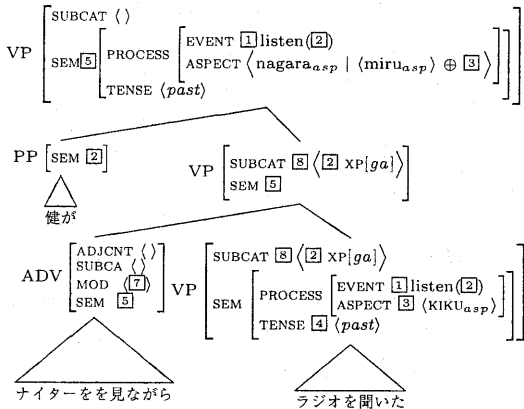
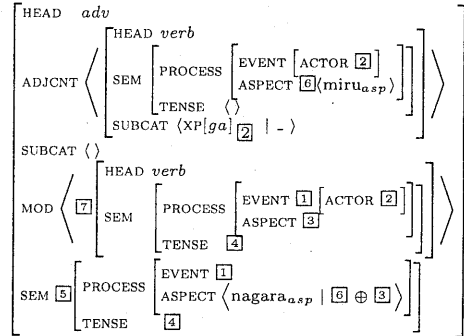


図 16: 接続助詞「ながら」の語彙項目 (上) と「健がナイターを見ながらラジオを聞いた」の構文木 (下)

### 5.1 「ながら」をめぐる諸現象についての考察

図 16(上) に「ながら」の語彙項目を示す。<sup>8</sup> 2.1 節で述べたように、その従属節は主格名詞句を伴えないが、この性質は、前接する動詞句は SUBCAT に *ga* 格名詞句を持つ、と隣接素性 (ADJ(A)C(E)NT) が制約している。従属節で表出できない主格名詞句は、主節の主語と同じでなければならないが、このことは変数 (2) の共有によって記述される。

図 16(下) に解析例を示す。「[健がテレビを見]ながら…」という解析が排除されることに注意されたい。これは「見」が主格名詞をとってしまうと、「ながら」に前接することができなくなるからである。

「ながら」には、前接する動詞句は時制を示せない、という語彙固有の制約がある。このことも、ADJCNT 中の TENSE は空でなければならない、というように局所的にとらえることができる。

<sup>8</sup> ⊕ は従属節がすでにもっている相と、主節のもつ相の「合成」を表す。「ながら」に関してはリストの単純な接続 (append) でよいと思われるが、他の語彙に関しては検討を要する。

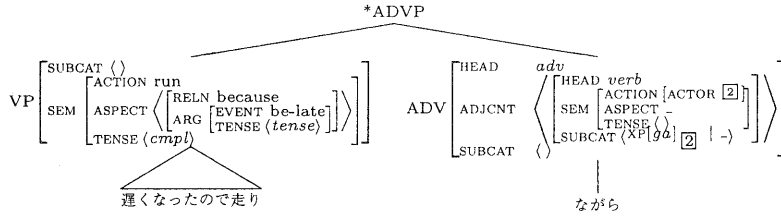


図 18: \*「遅くなったので走り」ながら (支度した) の構文木

## 5.2 「ながら」との対比としての「ので」

図 17 に接続助詞「ので」の語彙項目を示す。

「ので」は時制をもつ句に接続できるため、ADJUNCT 中の TENSE には、なんらかの時制をもつ、と指定される。それは主節の時制と合成され、文全体の時制を規定する。従属節と主節は独立の時制がとれるため、時制の計算はリストの単純な接続ではないが、重要なのは合成の結果が空ではない (i.e. *nelist(tense)* 型をもつ) という点である。

また、SUBCAT に関する制約はなく、主格名詞句を構成要素としてもつことができるので、主節の主語との関係も特に制約されない。

## 5.3 「ながら」と「ので」共起制約

先の接続助詞が共起できない例を図 18 に示す。

図 18 では、時制をもった節が「走り」に係ること、この節自身も時制をもつことになるため、「ながら」の制約に違反するのである。

南や吉本では「A 類節『走りながら』は B 類節『遅くなったので』を含むことはできない」という説明がされていたが、このようにみれば A, B, ... などの恣意的なラベルは不要であることがわかる。

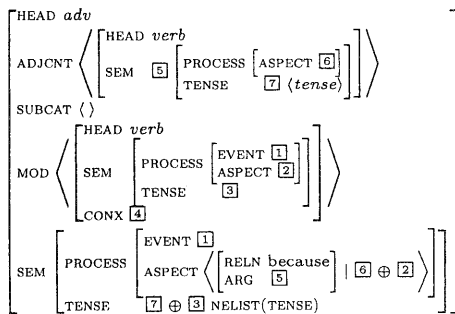


図 17: 接続助詞「ので」の語彙項目

## 6 おわりに

本論文では、日本語の性質を明らかにしつつ、実装上の問題も配慮しながら文法の形式化を試みた。処理における局所性や、統語や意味などの言語情報の性質をより重視した結果、従来の言語処理に用いられてきた文法よりも見通しがよい設計となった。より大きなデータを解析しつつ、文法を調整するといった課題は残っているが、一貫した統合処理で、これまで別処理に委ねられていた現象の扱いを体系的に関連付けることが可能となった。

## 参考文献

- [1] 郡司隆男. 日本語句構造文法に基づく効率的な構文解析の研究. 平成 5 年度文部省科学研究費補助金 (一般研究 (B)) 研究成果報告書, 大阪大学, 1994.
- [2] 中川裕志, 今仁生美, 郡司隆男, 田窪行則, 原田康成, 森辰則. 「ながら」について議論しながら. 日本認知科学会第 10 回大会論文集, pp. 40-41, 1993.
- [3] 大谷朗, 宮田高志, 松本裕治. HPSG にもとづく実用日本語文法について. 自然言語処理, Vol. 7, No. 5, pp. 19-39, 11 2000.
- [4] Carl J. Pollard and Ivan A. Sag. *Head-Driven Phrase Structure Grammar*. The University of Chicago Press, 1994.
- [5] James Pustejovsky. *The Generative Lexicon*. MIT Press, 1995.
- [6] Ivan A. Sag and Thomas Wasow. *Syntactic Theory: A Formal Introduction*. CSLI Publications, 1999.
- [7] 田窪行則. 統語構造と文脈情報. 日本語学, Vol. 6, No. 5, pp. 37-48, 5 1987.
- [8] 吉本啓. 日本語の文階層構造と主題・焦点・時制. 言語研究, No. 103, pp. 141-166, 1993.
- [9] Kei Yoshimoto, Chidori Nakamura, and Yoshiki Mori. A unified approach to tense in Japanese. In Akira Ikeya and Masahito Kawamori, editors, *PACLIC 14*, pp. 389-400, 2000.
- [10] 山田孝雄. 日本文法学概論. 室文館, 1936.
- [11] 金田一春彦 (編). 日本語動詞のアスペクト. むぎ書房, 1976.
- [12] 南不二夫. 現代日本語の構造. 大修館, 1974.
- [13] 南不二夫. 現代日本語の輪郭. 大修館, 1993.