

# NAIST JPSGにおける複合述語形成の形式化

大谷 朗

奈良先端科学技術大学院大学/大阪学院大学  
ohtani@utc.osaka-gu.ac.jp

松本 裕治

奈良先端科学技術大学院大学  
matsu@is.aist-nara.ac.jp

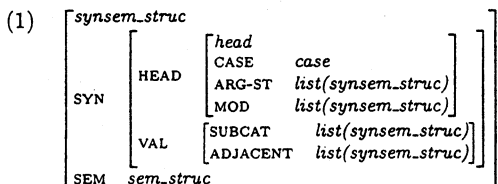
## 1 はじめに

日本語に特徴的な複合語の形式化は言語の諸相を見据えた宣言的な文法の記述や文解析には必須である。そのような言語(処理)観から、我々は単一化に基づいたNAIST日本語句構造文法(Japanese Phrase Structure Grammar: JPSG)<sup>1</sup> [3, 4] を基盤として様々な複合述語形成の形式化を試みた。

## 2 NAIST JPSGの概要

### 2.1 単純化した統語構造: NAIST JPSG/Phase 1

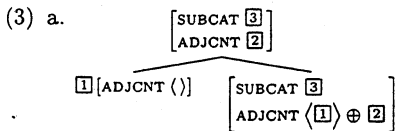
NAIST JPSG/Phase 1 [3] は当面の記述対象を日本語に限定したものの、HPSG [5, 7] に準じた「言語理論寄り」の文法モデルである。普遍原理 ( $P_1, \dots, P_n$ )、個別言語原理 ( $P_{n+1}, \dots, P_{n+m}$ )、選言的スキーマ ( $S_1, \dots, S_p$ )、および(1)のような素性構造<sup>2</sup>を持つ語彙情報 ( $L_1, \dots, L_q$ ) の単一化に基づき、日本語に特徴的な言語事実の集合、すなわち日本語を(2)のように形式化する。



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

### 2.2 隣接素性原理とPLRスキーマ

複合述語に関するPhase 1の解析戦略は、(3a)隣接素性原理と(3b)擬似語彙規則(Pseudo Lexical Rule: PLR)スキーマによる統語構造の単純化である。

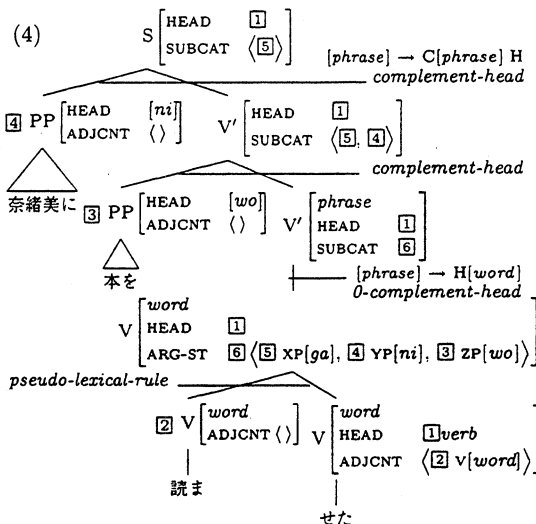


隣接素性 (ADJCNT) は下位範疇化の中でも特に隣接 (adjacent) 範疇を制約する。助詞や助動詞のように、他要素との隣接を強く要求する語が日本語には存在するが、

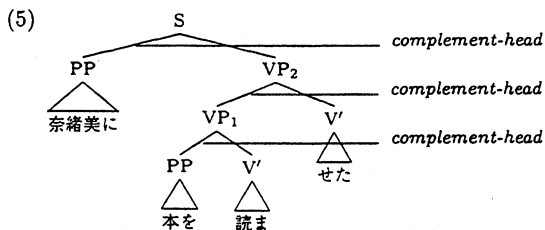
<sup>1</sup> ICOTによるJPSG[1]と区別してこのようによぶ。

<sup>2</sup> イタリック体による記述は部分情報構造の型名。型同士は上位・下位関係に関して束を成す。list( $\alpha$ )は $\alpha$ 型の素性のリスト。synsem\_structは語または句の持つ情報を記述するための型の総称であり、phrase, word, lexeme, ... といった下位型を持つ [7]。主辞 (HEAD) 素性の値 head は品詞を表し、noun, verb, ptcl, ... などの下位型を持つ。格 (CASE) 素性の値 case は格に関する情報。語や句・文の意味内容を担う意味 (SEM) 素性およびその値は [5] の記法を踏襲する。

例えば使役構文「健が奈緒美に本を読ませた」における複合述語「読ませた」は(4)のように分析される。<sup>3</sup>



(3b)を導入することの利点は、助動詞のADJCNT素性の指定によって姉妹をwordに制限することで、解析における複合述語句の曖昧性を抑制できる点である。例えば、(4)の助動詞サセのADJCNT素性が指定するタイプをwordに制限しないと(5)のようにも分析できる。

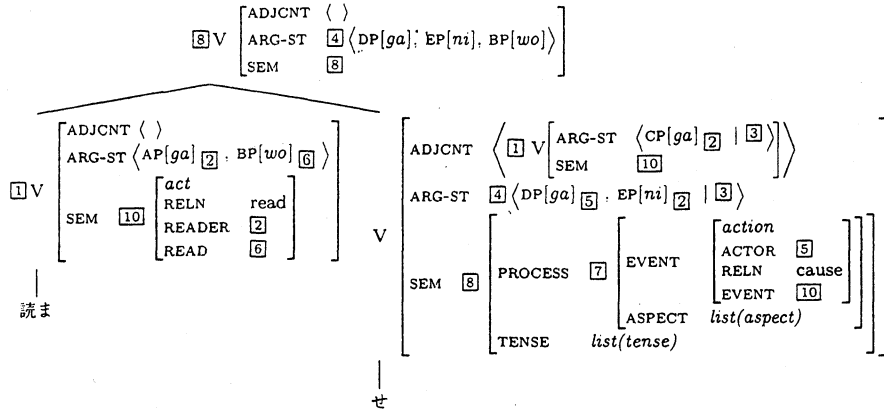


しかし、(5)は「奈緒美に」と「本を」の語順転換の解析が困難であり、またそれらの間に付加語が介在するとVP<sub>1</sub>とVP<sub>2</sub>のどちらに付加するかで曖昧性が増大する。一方(4)では解析木自身が語順転換や要素の介在・省略を直接的に反映するので、こうした問題は生じない。

Phase 1は、無効な解析木生成や曖昧性抑制といった実装上の問題を考慮してHPSGを独自に拡張したが、「特殊な」現象の処理を指向した「特殊な」計算機構は導入せず、またそうして統語解析に必要な文法の設計を単純化することでシステムを見通し良くしようとした。

<sup>3</sup> 構文木の構造は、いわゆる構成素の階層関係は反映していないが、そのような情報はSEM素性の構造が表現している。2.3節参照。

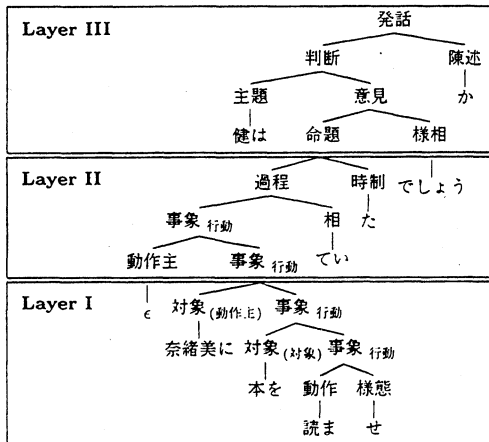
(9)



### 2.3 階層化した意味構造: NAIST JPSG/Phase 2

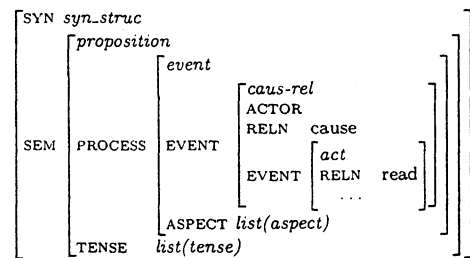
豊かな文末表現や修飾構造からも明らかなように、日本語文構造の一般的制約は統語情報よりも意味情報が寄与する。意味的側面の階層分析 [1] 自体は新しくないが、Phase 2[4] はそうした試みを (6) に基づく統語・意味 (・談話) 情報間の局所的制約として記述した。

(6)



言語情報の回合点は巨視的な文構造に反映されるという従来の主張に対し、(6) の意味タイプは、対応する統語構造を構成する主辞情報から構成的に計算される言語的対象として (7) の素性に基いて形式化される。

(7)



### 2.4 Layer: 統語・意味情報の統合構造モデル

語彙記載項目の情報構造に関する考え方は様々だが、統語・意味・形態が必須情報であることは概ね一致する。Phase 1 は統語情報を中心に文法記述の調整を試みたが、理論的に健全で計算的に有用なリソースを提供するという目標の達成には意味情報の記述が十分ではなかった。

Phase 2 は (6) の意味タイプの階層構成に基づき、(8) の方針で統語・意味情報の統合構造モデルを形式化した。

- (8) a. 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。
- b. 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。
- c. 制約は局所的に規定される。

局所的に制約された統語情報に基づく文構造の構築に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もそうした構築がなされる。(8) の方針は、(1) の素性構造を持つ計算辞書に対し、(7) の意味素性の拡充を中心に、またそのために必要な制約を (2) の体系に追加することで Phase 1 の拡張を図ることにもなる。

計算辞書をめぐるこうした理論的デバイスが、言語学だけでなく計算機科学において、どの程度妥当であり有用であるかは、可能な限り完全な辞書で実現しようとする事で検証される。Layer とは、文構造に写像された (6) のような述語の意味タイプを区別することで、複合述語形成を含む日本語に特徴的な事実を、よりの確に捉えるために導入した考え方である。

使役や受動などの Layer I に現れる様態辞の主な機能は、複合述語形成における意味情報の合成および統語構造の構成に連携した項構造の変化に他ならない。例えば (4) の使役構文の部分的解析木と sem\_struc の主要部分は (9) のように記述できる。Layer I は述語の語彙的意味とそれに必要な項構造の具現によって文の意味の中核を成すが、(7) は項構造を介して下位範疇化情報と連携し、述語が表わす事象とその参加者を制約するのである。

### 3 レイヤーに基づく複合述語形成

(6) の Layer をもとに、従来の研究が捉えようとした文の意味的性質を整理すると概ね (10) のようになる。

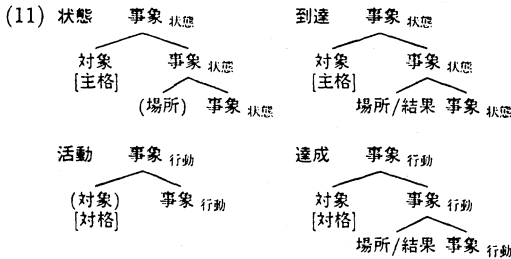
- (10) I. 対象や様態が表す動作などの描叙的機能 (事象)
- II. 描叙的な意味内容を相・時制などといった点から認定する判断的機能 (命題)
- III. 命題内容の陳述などといった談話的機能 (発話)

各 Layer の意味情報は (10) の関心が示すように性質を異にするが、単一化により情報は Layer 間を伝播し、文全体の意味の構成的計算に寄与する。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る。

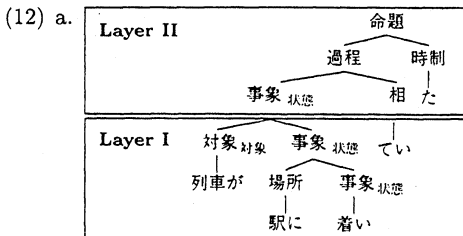
Layer III はさらに Layer II が規定した命題に対する話者の意見や判断といった発話内容を様々な様相・陳述辞と関連付けて計算する。しかしながら、そうした文脈情報への語彙の意味の寄与は少ないと考え、ひとまず Layer I, II が規定する (7) の *sem\_struct* とは区別する。<sup>4</sup>

#### 3.1 Layer I: 述語的語彙情報と語彙の複合

Layer I は項構造を介して下位範疇化情報と連携し、述語が表わす事象とその参加者を制約する。項構造は各述語に特有なものとして記載されるべき情報であるが、述語の数の膨大さに比べるとタイプは非常に限られる。項構造に関する述語タイプが限られるのは、(11) のみが意味的に妥当なプリミティブであることによる。



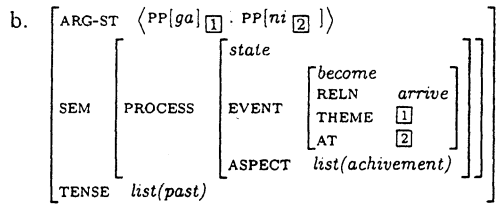
(11) は項構造の単なる視座ではなく、状態・到達・活動・達成といった述語の内在相の分類 [8] に対応する。<sup>5</sup>



<sup>4</sup> 文脈 (CON)(TE)(X)(T) 素性として記述することが考えられる。

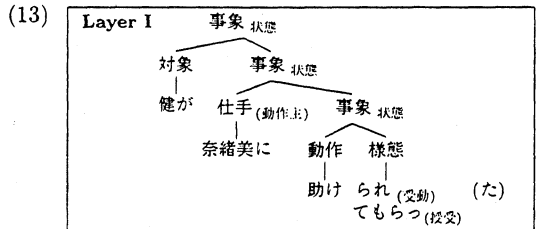
- (i)  $\left[ \begin{array}{l} \text{SEM} \quad \text{sem\_struc} \\ \text{CONX} \quad \text{conz} \end{array} \right]$

<sup>5</sup> テイル形や時を表す副詞との共起関係といった統語現象の説明においても (11) の分類は重要である。[9] を参照。



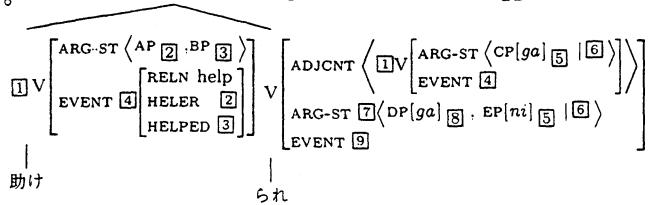
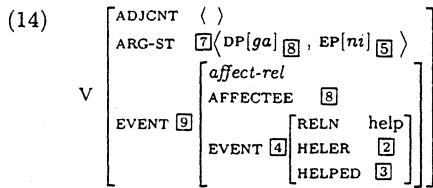
(12) に到達述語「着く」の Layer および意味素性を示す。詳しくは 3.2 節で見るが、単独の統語範疇である述語も、その意味情報は Layer I, II に分離された意味素性の合成であり、一種の抽象的な複合述語として扱われる。<sup>6</sup>

使役構文同様に受動・授受といった述語の様態変化を伴う文の構築は、Layer I における複合述語形成が負う。



いわゆる直接受動では、仕手・受手のどちらを主体に事象を描叙したのかという点から、事象の参加者に対応した項が担っている文法関係を語彙規則が変更し、一方対応する能動文のない間接受動では、事象に参加しない受影者と、影響を及ぼしている事象に対応した述語句を様態述語が下位範疇化するという分析が見受けられる。

それに対し (13) の部分的素性 (14) では Phase 1 の方針から (5) のような補文構造は仮定しない。そのような階層は Phase 2 の意味構造が反映し、直接・間接受動の違いもまた EVENT 素性を軸とした意味解釈が決定する。



<sup>6</sup> このような抽象化を推し進めれば、(11) のプリミティブの意味素性に以下のような概念構造的記述を導入することができる。

- 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>]]]  
 NAIST JPSG では生成辞書 [6] 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。

二種類の分析は、単一の統語構造に写像された意味構造の違いに還元されるが、詳細は別稿 [2] に委ねておく。<sup>7</sup>

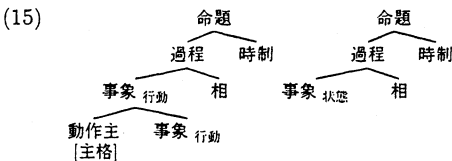
テモラウ形授受構文は受動との並行性が指摘されているが、NAIST JPSG は受動同様の統語情報を仮定する。述語が含意する意味は異なるが、両構文は参与者と主動詞が含意する事象の関係が同じであることによる。授受のみ主動詞がテ形であるが、それはモラウの隣接素性にそのような形態情報の制約あるからである。授受には統語構造の異なるテヤル・テクレル形もあるが、それらも利益関係 *benefit-rel* といった意味素性の下で授受構文のパラダイムを成している。

Layer I の複合述語形成には、いわゆる語彙的複合も含まれる。意味的制約が厳しいその種の複合では、述語の意味が元来の動詞の意味の単なる合成とは限らない。そのような操作の形式化には述語の意味をより詳細に記述しつつ、特に派生的意味を含意するものについては生成辞書 [6] のような枠組の導入を検討しなければならないが、今後の課題となっている。

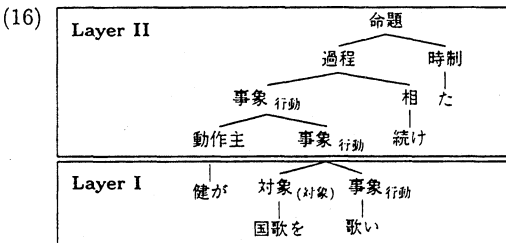
### 3.2 Layer II: 指定的語彙情報と統語的複合

Layer II の機能は、Layer I で参与者が規定した事象を様々な意味的側面について認定することにある。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る。

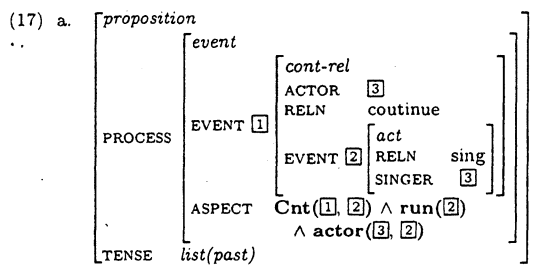
主格名詞は、意思的事象では時制を明示する節に動作主として現れるのに対し、非意思的事象では時制の区別がない節にも対象として現れるが、このことは時制に関する Layer II のプリミティブ (15) および Layer I の (11) による抽象的複合述語を仮定することで説明される。



Layer II の複合述語形成には、いわゆる統語的複合がある。任意のタイプの主動詞と複合が可能で、生産性も高い、そうした述語の代表的な意味タイプが相である。相助動詞はコントロール・繰り上げ述語の一方、あるいは両方の統語的特性を持つが、NAIST JPSG は Phase 1 の方針上補文構造をとらないため、これらを区別しない。



<sup>7</sup> 間接受動および直接受動のうち仕手がニ格を取るタイプに限る。ニヨッテでマークされる直接受動は、英語の *by*-phrase 受動文同様に語彙規則を導入して動作主を降格しておく必要がある。[2] 参照。



- b. *sing*(*e*) : *e* は歌うという動作もしくは事象  
*actor*( $\alpha$ , *e*) :  $\alpha$  は *e* の動作の主体  
*Cnt*(*e*, *e'*) : *e* は「ある期間事象 *e'* が起きた」という事象

(17a) では、Phase 2 に基づき ASPECT および TENSE の値をリストとしている。これは、動詞本来の情報が相助動詞などの Layer II の述語を伴うことで、それらの情報についてさらに指定を受けることを記述した。ツツケルやテイル形の相解釈は継続と結果・反復に大別されるが、(17a) は反復を表わしている。ただし、語彙に固有な相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、この点は後考を期したい。

### 4 おわりに

NAIST JPSG の複合述語形成は、語や句という大きな言語対象に依拠する伝統的区分では成し得なかった述語内の細分類・述語間の共通性について、統語・意味・形態情報に関する単純な素性と素性間の僅かな制約だけで規定できた見通しの良い分析および実装となった。もちろん理論的に興味深い文法が記述できても、そうした分析の正しさまでもは保証できないが、実装することで特定の分析的枠組の限界を示すことは十分可能だろう。

### 参考文献

- [1] 郡司隆男. 日本語構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] A. Ohtani. An HPSG account of Japanese passive. In *Proceedings of the Second ICCS and the 16th Annual Meeting of the JCSS Joint Conference*, pp. 527-530. The Japanese Cognitive Science Society, 1999.
- [3] 大谷, 宮田, 松本. 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] 大谷, 宮田, 松本. 述語の隣接と述語の付加における意味的階層性. *情報処理学会研究報告*, 第 140 (2000-NL-140) 巻, pp. 103-110. 情報処理学会, 2000.
- [5] C. J. Pollard and I. A. Sag. *Head-Driven Phrase Structure Grammar*. The University of Chicago Press, 1994.
- [6] J. Pustejovsky. *The Generative Lexicon*. MIT Press, 1995.
- [7] I. A. Sag and T. Wasow. *Syntactic Theory: A Formal Introduction*, Vol. 92 of *CSLI Lecture Notes Series*. CSLI Publications, Stanford, California, 1999.
- [8] Z. Vendler. *Linguistics in Philosophy*. Cornell University Press, 1967.
- [9] 金田一春彦. *日本語動詞のアスペクト*. むぎ書房, 1976.